The Scientific Journal of Business and Finance

https://caf.journals.ekb.eg

Calibration Weights Estimation in Stratified Random Sampling for Multiple Skewed Variables

Asmaa helmy [a], Ramadan Hamid[b], Lobna Eid [c], Abeer Elrefaey [d]

[a] Assistant Lecturer of Statistics Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

[b] Professor of StatisticsFaculty of Economics and Political Science, Cairo University, Egypt. Research Professor, Social Research Center, AUC.

[c] Professor of Statistics Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

[d] Assistant Professor of Statistics Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

Published online: **December 2024.**

To cite this article: helmy, Asmaa, Hamid, Ramadan, Eid, Lobna & Elrefaey, Abeer. **Calibration Weights Estimation in Stratified Random Sampling for Multiple Skewed Variables**, The Scientific Journal of Business and Finance, 44, (4),182-195

DOI: 10.21608/caf.2024.392545

*Corresponding author: lobnaalatyeb@azhar.edu.eg,

Calibration Weights Estimation in Stratified Random Sampling for Multiple Skewed Variables

Asmaa helmy

Assistant Lecturer of Statistics Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

Ramadan Hamid

Professor of StatisticsFaculty of Economics and Political Science, Cairo University, Egypt.

Research Professor, Social Research Center, AUC.

Lobna Eid

Professor of Statistics and Dean of Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

Abeer Elrefaey

Assistant Professor of Statistics Faculty of commerce, Al-Azhar University Girl's Branch, Cairo, Egypt.

Article History

Received 17 August2024, Accepted 7 November 2024, Available online December 2024.

Abstract

The calibration estimation has been applied in different fields and it constitutes a setup for many statistical problems. It has become an important topic in recent research on estimation in survey sampling. Calibration provides a systematic way to incorporate auxiliary information in the procedure. In this paper, using a *Goal Programming* (GP) model is newly suggested for the estimation of calibration weights in stratified random sample. This paper considers the skewed variables case with different calibration constraints as a result of using log-normal distribution for the study variable in stratified sample survey, and applied the suggested model to Egyptian Labor Market data. The results of the proposed approach for this survey are satisfying.

Keywords: Calibration Weights Estimation; Stratified Random Sampling; Goal Programming Model; Auxiliary Variables; Log-Normal Distribution.

1. Introduction:

A common method for modifying the initial weights under auxiliary information is calibration estimation, which involves minimizing a distance measure that serves as the foundation for calibration constraints. In the literature, researchers structure new calibration weights in stratified sampling to improve the precision of population parameter estimations. The process of creating new calibration weights consists of two main components: a distance measure and a set of calibration constraints [Ozgul, (2019)].

In finite population sampling, there are various methods for using auxiliary information to improve the precision of the population parameter estimators. More effective estimators for population totals and means can be obtained in several ways. Any estimating technique needs to first the weight sample data before calculating weighted averages. Calibration is used for changing sample weights in order to improve estimation, as these weights need to be adjusted frequently to produce better estimates [Nidhi *et al.* (2016)].

When making statistical estimates, researchers often rely on the assumption that relevant variables follow a normal distribution, an assumption that can sometimes be incorrect. There are many variables that follow skewed distributions. It has been proven by many types of research that estimating data that follows skewed distributions, and has been assumed to be based on a normal distribution, can lead to many problems, including the problem of bias. In this paper, we will estimate the calibration weights under skewed distributions, by using the *Goal Programming* (GP) approach as a flexible method for solving this problem.

The rest of this article is organized as follows: Section 2 is about previous studies of calibration estimation. In section 3, the calibration weights estimation approach for stratified random sample is presented. The MP model for calibration weights estimation of the optimum strata is given in section 4. The suggested GP approach for calibration weights estimation of the optimum strata is reviewed in section 5. Section 6 is presented suggested GP approach for calibration weights estimation of the optimum strata with skewed distribution. In section 7, the application of the suggested model is conducted using the Egyptian Labor Market sample survey. And present the result of this application in section 8. Finally, the conclusions introduced in section 9.

2. Previous Studies of the Calibration Estimation:

Numerous survey statisticians have explored calibration estimation. The first study to use calibration estimates in survey sampling was by [Deville and Sarndal (1992)]. The calibration requirements were first established when auxiliary variables were included. Deville and Särndal first used the phrase "calibration estimation" to describe a process of minimizing a distance measure between design and calibrated weights that is subject to calibration equations.

The first work to extend the calibration approach to the stratified random sampling (STRS) design was [Singh *et al.* (1998)], who also established the calibration approach for the combined Generalized Regression Estimator (GRE). Many authors have contributed to the theory of calibration estimation in STRS after Singh *et al.* (1998). Singh (2003), they obtained calibration estimator of the population mean $\bar{Y}$ under stratified sampling given by

$$\bar{Y}_{st}^{\ s} = \sum_{h=1}^{L} W_h^* \ \bar{y}_h. \qquad (2.1)$$

Here W_h^* ; $h = 1, 2, \dots, L$ are calibrated weights which are chosen to minimize the sum of the chi-square distance measure given in

$$\sum_{h=1}^{L} \frac{\left(W_h^* - W_h\right)^2}{Q_h W_h} \qquad (2.2)$$

Using *Lagrange Multiplier Technique* (LMT) subject to two fundamental calibration constraints,

$$\sum_{h=1}^{L} W_h^* \bar{x}_h = \bar{X}, \tag{2.3}$$

$$\sum_{h=1}^{L} W_h^* = 1. \tag{2.4}$$

Where $\bar{x}_h$ and $\bar{X}$ are h^{th} stratum sample mean and population mean respectively of the auxiliary variable X. Q_h is a carefully chosen constant that determines the final form of the calibrated estimators to decide different forms of estimators, [Berrada *et al.* (1996)], It can be equal to $\frac{1}{\bar{x}_h}$.

[Tracy *et al*. (2003)], also used LMT according to two restrictions to minimize the function in (2.2) and obtain the calibration estimator of the population mean $\bar{Y}$ under stratified sampling. However, it differs from Singh in (2.4) in the second set of constraints, where the auxiliary variable's second-order moments were used. Thus, the following formulation of the second constraint can be made:

$$\sum_{h=1}^{L} W_h^* s_{xh}^2 = \sum_{h=1}^{L} W_h S_{xh}^2 , \tag{2.5}$$

where s_{xh}^2 and S_{xh}^2 are the sample and population variance of the auxiliary variable in the h^{th} stratum respectively.

[Kim *et al.* (2007)], different calibration approach ratio estimators were proposed and the variance estimator of the calibration approach ratio type estimator in stratified sampling for four estimators was produced. [Rao *et al.* (2012)], developed new calibration weights for several auxiliary variables by applying LMT to minimize the chi square distance measure in (2.2), subject to the following two calibration restrictions:

$$\sum_{h=1}^{L} W_h^* \bar{x}_h = \bar{X}_1, \tag{2.6}$$

$$\sum_{h=1}^{L} W_h^* \bar{x}_{2h} = \bar{X}_2. \tag{2.7}$$

[Sud *et al*. (2014)] have developed a regression-type estimator of population total using the calibration approach suggested by [Deville and Särndal (1992)], When there is a negative correlation between the study and the auxiliary variables. Moreover, [Koyuncu and Khadilar, (2016)], used by minimizing the chi-square distance measure (2.2) and applying LMT to their calibration estimator, subject to three constraints that are the same as the constraints proposed in Singh, (2003) in (2.3), (2.4), and Tracy, (2003) in (2.5). They considered the suggested using them at one optimization problem simultaneously.

Clement (2015), created a modified population mean calibration estimator under the new restrictions by utilizing the auxiliary variable's mean, variance, and coefficient of the skewness. calibration estimator of the population mean $\bar{Y}$ under stratified sampling is given by:

$$\bar{Y}_{st}^{\ c} = \sum_{h=1}^{L} W_h^* \bar{y}_h. \tag{2.8}$$

Clement using are calibrated weights which are chosen to minimize the sum of the chi-square distance measure given by (2.2) subject to the three calibration restrictions that follow:

$$\sum_{h=1}^{L} W_h^* \bar{x}_h = \sum_{h=1}^{L} W_h \bar{x}_h^*, \quad (2.9)$$

$$\sum_{h=1}^{L} W_h^* s_{xh}^2 = \sum_{h=1}^{L} W_h s_{xh}^{*2}, \quad (2.10)$$

$$\sum_{h=1}^{L} W_h^* \hat{\beta}_{1xh} = \sum_{h=1}^{L} W_h \beta_{1xh}^*. \quad (2.11)$$

Where, $\bar{x}_h$ & s_{xh}^2 are first phase of the samples mean and variance of the auxiliary variable X, respectively, $\bar{x}_h^*$ & s_{xh}^{*2} are the second phase of the samples mean and variance of the auxiliary variable X, respectively, and β_{1xh}^* & $\hat{\beta}_{1xh}$ are the auxiliary variable X's sample coefficient of skewness in the first and second phase samples.

Rabee *et al*. (2021), developed the calibration estimation under STRS by incorporating two auxiliary variables, and formulated as the *Goal Programming* (GP) Model. They used the GP approach for solving different optimization problems. That contains too restrictive hard constraints.

Most authors of the calibration estimation literature looked at the calibration weights estimation as an optimization problem using the *Lagrange Multiplier Technique* (LMT), wherein the auxiliary variable-related constraints are minimized with the Chi-square distance function estimated. In the end, the optimal calibrated weights were estimated by applying the (LMT). Because the *Lagrange Multiplier Technique* requires the need that all model equations be differentiable functions, but there are often equations that cannot be differentiated. Therefore, it is important to explore an alternative method that can provide higher flexibility when dealing with different optimization problems. Hence, it is recommended to use MP technology as an effective solution to the problem of estimating calibration weights.

3. Calibration Weights Estimation Approach for Stratified Random Sample:

In sample surveys, sample weights are frequently adjusted through a process called calibration. We treat the calibration as a problem of optimization and show that the calibrated weights depend on the optimization function selected. Suppose the population consists of h strata with N_h units in the h^{th} stratum from which a simple random sample of size n_h is taken without replacement. Let total population size be $N = \sum_{h=1}^{L} N_h$ and sample size be $n = \sum_{h=1}^{L} n_h$, respectively. For the h^{th} stratum, let $W_h = \frac{N_h}{N}$ be the stratum weights, and $\bar{y}_h$, $\bar{Y}_h$ are the sample and population means, respectively, for the study variable. It must be mentioned that under the STRS design, the unbiased estimator of the population mean:

$$\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h, \quad (3.1)$$

is given by [Cochran, (1977)]

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h, \quad (3.2)$$

where $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$, (3.3)

denotes the stratum sample mean h^{th}. Furthermore, the $\bar{y}_{st}$ estimated variance under the *Simple Random Sample Without Replacement* (SRSWOR) scheme is given by:

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) W_h^2 \, S_{hy}^2$$

$$= \sum_{h=1}^{L} \frac{W_h^2 \, S_{hy}^2}{n_h}, \quad (3.4)$$

where $s^2_{hy} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$ (3.5)

is the sample variance of Y in the stratum $h, h = 1, 2, \dots, L$. The usual estimator of population means that given (2.1) is unknown. Assume that X_{ijh} denotes the i^h unit's value of the j^{th} auxiliary variable in the stratum h; $h = 1, 2, \dots, L$, $i = 1, 2, \dots n_h$ & $j = 1, 2, \dots, P$. The estimate of population means $\bar{X}_j = \sum_{h=1}^{L} W_h \, \bar{X}_{jh}$ are accurately known by using the auxiliary information X_j.

Assume additionally that there are several auxiliary variables and that the sample values for these variables are known, either precisely or informally. The main objective of the calibration problem is to discover new weights W_h^* that take into account the auxiliary data in order to improve the initial weights.

In the presence of the population's parameter through merging data from more than one auxiliary variable, the suggested calibrated estimator of the population mean $\bar{Y}$ under STRS is given by

$$\bar{y}_{st}^{c} = \sum_{h=1}^{L} W_h^* \, \bar{y}_h. \quad (3.6)$$

where new weights W_h^*; $h = 1, 2, \dots, L$ represents the estimated calibration weights. When multiple auxiliary variables X_j, $j = 1, 2, \dots, P$ are available, the calibration weights estimation W_h^* are so chosen such that the sum of the chi-square type distances given (2.2) in minimum, subject to some specific calibration constraints, as following [(Rao *et al*, (2012)]

$$\bar{X}_j = \sum_{h=1}^{L} W_h^* \bar{X}_{jh}, \quad j = 1, 2, \dots, P. \quad (3.7)$$

Note that, the estimated variance of calibration estimation $\bar{y}_{st}^{c}$ scheme is given by:

$$\hat{V}(\bar{y}_{st}^{c}) = \sum_{h=1}^{L} \left(\frac{1-f_h}{n_h}\right) W_h^{*2} \, s_{hy}^2, \quad (3.8)$$

where s_{hy}^2 is the sample variance of Y in the stratum $h, h = 1, 2, \dots, L$. The primary goal of this study is to introduce a new set of calibrated weights in order to offer new Multivariate

Calibration Estimation Incorporating two auxiliary variables using these calibrated weights will increase the precision of the results for stratified random sampling.

4. MP Model for Calibration Weights Estimation of the Optimum Strata:

In this section, the optimal calibrated weights estimation problem is formulated as the MP model. It is known that any MP model consists of an objective function and constraints of the decision variable. The objective function that used in almost the previous studies for calibration weights estimation in stratified random sample is minimizing the chi-squared distance function. In this research, we used Manhattan distance (L_1 Norm) from the population stratum, Rabee (2022), using L_1 decreases the impact of the outlier's existence that the design weights W_h can be contains it.

The calibrated weights W_h^*; $h = 1, 2, \dots, L$, and the design weights (W_h; $h = 1, 2, \dots, L$) are represented by a Manhattan distance measure in the objective function. Thus, it can be stated as follows:

$$\text{Minimize } Z = \sum_{h=1}^{L} |W_h - W_h^*| \qquad (4.1)$$

where W_h and W_h^* are the design weight and calibrated weight for the h^{th} stratum respectively; $h = 1, 2, \dots, L$.

The calibration weights W_h^* are chosen to minimizing a given distance measure subject to satisfying constraints related with auxiliary variables. The three constraints include the main key of the calibration procedure. The constraints can be stated as follows:

$$\sum_{h=1}^{L} W_h^* \bar{x}_{jh} = \bar{X}_j, \qquad (4.2)$$
$$\sum_{h=1}^{L} W_h^* s_{xjh}^2 = S_{Xj}^2, \; j = 1, 2, \dots, P. \qquad (4.3)$$
$$\sum_{h=1}^{L} W_h^* = 1. \qquad (4.4)$$

Where $\bar{x}_{hj} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{ijh}$ and $\bar{X}_j$ are h^{th} stratum sample mean and population mean respectively of the j^{th} auxiliary variable in eq (4.2), and where $s_{xjh}^2 \left\{\sum_{i=1}^{n_h} \left(x_{ij} - \bar{x}_{jh}\right)^2 / (n_h - 1)\right\}$ & $S_{Xj}^2 \left\{\sum_{i=1}^{N} \left(X_{ij} - \bar{X}_j\right)^2 / (N-1)\right\}$ are the sample and population variances of the auxiliary variable X, respectively that in eq (4.3).

In addition, two other constraints are suggested to improve the precision of the calibration weights estimation. Can be expressed those constraints as following:

$$\sum_{h=1}^{L} \frac{W_h^{*2} S_{hy}^2}{n_h} \leq v \qquad (4.5)$$
$$\left|\sum_{h=1}^{L} W_h^* \bar{y}_h - \bar{Y}\right| \leq \epsilon \qquad (4.6)$$

In eq (4.5) $\sum_{h=1}^{L} \frac{W_h^{*2}\, s_{hy}^2}{n_h}$ represents the estimated variance of $\bar{y}_{st}$, where $s^2_{hy} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$ is the sample variance of the study variable Y_h in the stratum h^{th}, $h = 1, 2, \dots, L$.

The constraint in eq (4.6) was added in order to keep the estimate unbiased of the calibration weights estimator [Rabee *et al*. (2021)].

Then to determining the optimum calibration weights estimation can be formulated using MP model as follows:

Find W_h^*; $h = 1,2, \dots, L$, that

Minimize $Z = \sum_{h=1}^{L} |W_h - W_h^*|$, (4.7)

Subject to

$\sum_{h=1}^{L} W_h^*\, \bar{x}_{hj} = \bar{X}_j, \quad j = 1, 2, \dots, P,$ (4.8)

$\sum_{h=1}^{L} W_h^*\, s_{xjh}^2 = S_{xj}^2, \quad j = 1, 2, \dots, P,$ (4.9)

$\sum_{h=1}^{L} W_h^* = 1,$ (4.10)

$\sum_{h=1}^{L} \frac{W_h^{*2}\, s^2_{hy}}{n_h} \le v,$ (4.11)

$|\sum_{h=1}^{L} W_h^* \bar{y}_h - \bar{Y}| \le \epsilon,$ (4.12)

$W_h^* \ge 0, \; h = 1, 2, \dots, L.$ (4.13)

Where

W_h: the population weights of the stratum h^{th}, h is called number of the strata.

W_h^*: the population calibration weights of the stratum $h^h, h = 1, 2, \dots, L$.

$\bar{x}_{jh}$: the sample mean for the j^{th} auxiliary variable; $j = 1, 2, \dots, P$.

$\bar{X}_j$: the population mean for the j^{th} auxiliary variable.

s_{xjh}^2: the sample variance for the j^{th} auxiliary variable.

S_{Xj}^2 : the population variance for the j^{th} auxiliary variable.

s^2_{hy}: the sample variance of the study variable y for stratum h^{th}.

n_h: the sample size of the stratum h^{th}.

v : the estimated variance of the population mean for the calibrated weights estimator, which will be obtained from previous studies.

$\bar{y}_h$: the study variable's sample mean y_h for stratum h^{th}.

$\overline{Y}$: the population mean for the variable under study. That can be obtained it from previous studies.

ϵ: the positive constant with small value about the bias estimate's calibrated weights estimators, (given small value).

5. Suggested GP Approach for Calibration Weights Estimation of the Optimum Strata:

The GP approach has been accepted as a basic MP model for solving different optimization problems. The primary goal of this section is to introduce a new set of calibrated weights in order to offer new Multivariate Calibration Estimation Incorporating two auxiliary variables using these calibrated weights will increase the precision of the results for stratified random sampling. To estimate the calibration weights in stratified random sample using GP approach. That can be expressed as follows:

Find W_h^* , dn_i, dp_i; $h = 1,2,\dots,L,\ i = 1,\dots,k$ that

Minimize : $Z = \sum_i^k (dn_i + dp_i)$ (5.1)

Subject to

$$W_h^* + dn_1 - dp_1 = W_h, \quad h = 1,2,\dots,L \tag{5.2}$$

$$\sum_{h=1}^{L} \frac{W_h^{*2}\, s^2{}_{hy}}{n_h} + dn_2 - dp_2 = v, \tag{5.3}$$

$$\sum_{h=1}^{L} W_h^*\, \bar{x}_{hj} + dn_3 - dp_3 = \bar{X}_j, \tag{5.4}$$

$$\sum_{h=1}^{L} W_h^*\, s_{xhj}^2 + dn_4 - dp_4 = S_{xj}^2 , \tag{5.5}$$

$$\sum_{h=1}^{L} W_h^* \bar{y}_h + dn_5 - dp_5 = \bar{Y}, \tag{5.6}$$

$$\sum_{h=1}^{L} W_h^* = 1, \tag{5.7}$$

$$W_h^*,\ dn_i,\ dp_i \geq 0,\ h = 1,2,\dots,L,\ i = 1,\dots,k.$$

Where, dn_i, dp_i are negative and positive deviation variables respectively of the i^{th} goal , i is the total number of constraints, and $i = 1,\dots,k$. Also, the first constraint was added in this approach as follows: $W_h^* + dn_1 - dp_1 = W_h$, where it represents the main goal when using calibration estimation, which was previously expressed as minimize $Z = \sum_{h=1}^{L} |W_h - W_h^*|$, as it is equivalent to that expression in eq (5.1).

6. Suggested GP Approach for Calibration Weights Estimation of the Optimum Strata with Skewed Distribution:

In this section, the suggested the GP Approach for Calibration Weights Estimation of the Optimum Strata under skewed distributions is presented. Let the probability distribution for the study variable of Y is log-normal. The probability density function is as follows:

$$f(y) = \frac{1}{y\,\sigma\sqrt{2\pi}} \left[exp - \left(\frac{log(y) - \mu}{\sigma\sqrt{2}} \right) \right], \quad -\infty < \mu < \infty,\ \sigma \geq 0 \tag{6.1}$$

where the study variable is y and the parameters $-\infty < \mu < \infty,\ \sigma \geq 0$ all are real numbers.

For stratified random sample with the strata and using the GP approach to obtain the calibration weights estimation for Log-Normal distribution is study variable with two auxiliary variables the j^{th}; $j = 1, 2.$, the GP is as follows:

Find $W_h^*, dn_i, dp_i; h = 1,2, \dots, L,$ that

Minimize: $Z = \sum_i^k (dn_i + dp_i),$ (6.2)

Subject

$$W_h^* + dn_1 - dp_1 = W_h, \quad h = 1,2, \dots, L \quad (6.3)$$

$$\sum_{h=1}^{L} \frac{W_h^{*2}}{n_h} \left[e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)\right] + dn_2 - dp_2 = v, \quad (6.4)$$

$$\sum_{h=1}^{L} W_h^* \bar{x}_{1h} + dn_3 - dp_3 = \bar{X}_1, \quad (6.5)$$

$$\sum_{h=1}^{L} W_h^* \bar{x}_{2h} + dn_4 - dp_4 = \bar{X}_2, \quad (6.6)$$

$$\sum_{h=1}^{L} W_h^* s_{x1h}^2 + dn_5 - dp_5 = S_{X1}^2, \quad (6.7)$$

$$\sum_{h=1}^{L} W_h^* s_{x2h}^2 + dn_6 - dp_6 = S_{X2}^2, \quad (6.8)$$

$$\sum_{h=1}^{L} W_h^* \bar{y}_h + dn_7 - dp_7 = \bar{Y}, \quad (6.9)$$

$$\sum_{h=1}^{L} W_h^* = 1, \quad (6.10)$$

$$W_h^*, dn_i, dp_i \geq 0, \ h = 1,2, \dots, L, \ i = 1, \dots, k.$$

Where, dn_i, dp_i are negative and positive deviation variables respectively of the i^{th} goal, i is the total number of constraints, and $i = 1, \dots, k$ And where, $Var(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^{*2}}{n_h} \left[e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)\right]$ is the variance of stratified random sample mean, and $\left[e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)\right] = s^2_{hy}$ is the variance of stratum for variable in the h^{th} stratum under Log-Normal distribution.

7. The Application of the suggested Model:

In this section, the suggested model is applied on Egyptian Labor Market Survey, where the main study variable is the number of education years, and two auxiliary variables are age and the wealth index, In addition, there are two strata according to the gender variable (males and females). This application aims to use the suggested model of the calibration weights and find the variance of the calibrated estimator of numbers of the education years average. Tables (7.1) and (7.2) present the main information for the data of the considered variables for the application from the samples and the population.

Table 7.1: Basic information about the sample survey

Strata(h)	n_h	$\bar{x}_{1h}$	$\bar{x}_{2h}$	$\bar{y}_h$	s_{x1h}^2	s_{x2h}^2	s^2_{yh}
Male 1	9206	44.36	1.93	11.206	150.121	422.485	26.169
Female 2	6940	43.09	5.33	11.127	143.093	479.801	27.957
Total	16146						

Table 7.2: Basic information about the population

Strata (h)	Stratum size (N_h)	Population stratum weights (W_h)	Some parameter of population
Male 1	19179990	0.519127644	$\bar{X}_1 = 47.04$
Female 2	17766588	0.480872356	$\bar{X}_2 = -1.64$
	N=36946578		$S_{x1}^2 = 158.67$
			$S_{x2}^2 = 1732.807$

This section is concerned with applying the suggested calibration weights estimation with skewed distribution by using *Goal programming* (GP). When the probability distribution for the study variable of (Y) is the log-normal distribution and the application on Egyptian Labor Market Survey, the estimated value of the parameters of the education of years variable is $\mu = (2.2999, 2.2818)$, $\sigma^2 = (0.2331, 0.2551)$ for the strata with the mean is expected $(Y) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$, that equal to (11.206, 11.127) and the variance is $Var(Y) = [exp\ (\sigma^2) - 1]\ exp\ (2\mu + \sigma^2)$. That equal to (26.169, 27.957). μ is estimated from the data average ln (y_i) and σ^2 is estimated by the variance of ln (y_i)

For using the GP approach we will obtain the calibration weights estimation with Log-Normal distribution as follows:

Find W_h^*, dn_i, dp_i; $h = 1, 2.$ that

Minimize $Z = \sum_i^7 (dn_i + dp_i)$, (7.1)

Subject to

$W_1^* + dn_1 - dp_1 = 0.519,$ (7.2)

$W_2^* + dn_2 - dp_2 = 0.480,$ (7.3)

$26.169\ W_1^{*2} + 27.957\ W_2^{*2} + dn_3 - dp_3 = 0.00169758,$ (7.4)

$44.36\ W_1^* + 43.09\ W_2^* + dn_4 - dp_4 = 47.04,$ (7.5)

$1.93\ W_1^* + 5.33\ W_2^* + dn_5 - dp_5 = -1.64,$ (7.6)

$150.121\ W_1^* + 143.093\ W_2^* + dn_6 - dp_6 = 158.67,$ (7.7)

$422.485\ W_1^* + 479.801\ W_2^* + dn_7 - dp_7 = 1732.807,$ (7.8)

$W_1^* + W_2^* = 1,$ (7.9)

$W_h^* \geq 0, , h = 1, 2.$

Where, $W_1 = 0.519$, $W_2 = 0.480$, $dn_i, dp_i \geq 0$ & $h = 1, 2$ & $i = 1, 2 \dots, 7$.

In this application, we used the software GAMS (General Algebraic Modeling System) models to solve these models.

8. Result of the Application:

This section presents the results of solving the multivariate calibration weights estimation model proposed by equations (7.1) to (7.9), which is used with application to Egyptian labor market data using GAMS and is solved by nonlinear programming. The following table (8.1) shows the estimated calibration weights given as follows.

Table 8.1: The Estimated Calibrated Stratum Weights

Strata (h)	weights	calibration weights
1	0.5191	0.519
2	0.4809	0.481
Sum	1	1

Through the values mentioned in the previous table and the results achieved, the values for estimating the multivariate calibration weights were obtained. Therefore, the estimated variance value for $\bar{y}_{st}$ is calculated as 0.001697693, while the estimated variance value based on the weights is 0.00169758. As a result, using multivariate calibration weights with skewed distribution improves the estimate's accuracy and gives the estimated variance is unbiased estimator for the population mean.

9. Conclusion:

In this paper, a new approach is proposed to estimate the multivariate calibration of the population mean of the studied variable with a skewed variable under the STRS model. By introducing additional calibration constraints for the lognormal distribution of the studied variable and applying the objective programming method, new calibration weights are generated. The accuracy of the proposed estimator is evaluated using this model applied to the Egyptian labor market data. According to the results, the proposed estimator is better than the previously used alternative calibration in terms of estimations. In particular, the GP technique provides a more flexible and efficient calibration estimator in the presence of a skewed variable for stratified sample design, in addition to the estimated variance is considered an unbiased estimator of the population mean.

References:

Berrada, I., Ferland, J. A., & Michelon, P. (1996). A multi-objective approach to nurse scheduling with both hard and soft constraints. *Socio-economic planning sciences*, *30*(3), 183-193.

Clement, E. P. (2015). Calibration approach separate ratio estimator for population mean in stratified sampling. *International Journal of Modern Mathematical Sciences*, *13*(4), 377-384.

Cochran, W. G. (1977). *Sampling techniques*. john wiley & sons.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, *87*(418), 376-382.

Kim, J. M., Sungur, E. A., & Heo, T. Y. (2007). Calibration approach estimators in stratified sampling. *Statistics & probability letters*, *77*(1), 99-103.

Koyuncu, N., & Kadılar, C. (2014). A new calibration estimator in stratified double sampling. *Hacettepe Journal of Mathematics and Statistics*, *43*(2), 337-346.

Nidhi, Sisodia, B. V. S., Singh, S., & Singh, S. K. (2017). Calibration approach estimation of the mean in stratified sampling and stratified double sampling. *Communications in Statistics-Theory and Methods*, *46*(10), 4932-4942.

Ozgul, N. (2019). New calibration estimator based on two auxiliary variables in stratified sampling. *Communications in Statistics-Theory and Methods*, *48*(6), 1481-1492.

Rabee, S., Hamed, R., Kassem, R., & Rashwaan, M. (2021). A goal programming approach for multivariate calibration weights estimation in stratified random sampling. *Mathematics and Statistics*, *9*(3), 326-334.

Rabee, S., Hamed, R., Kassem, R., & Rashwaan, M. (2022). A goal programming approach for generalized calibration weights estimation in stratified random sampling.

Rao, D., Khan, M. G., & Khan, S. (2012). Mathematical programming on multivariate calibration estimation in stratified sampling. *World academy of science, engineering and technology*, *72*, 58-62**.**

Singh, S., Horn, S., Yu, F. (1998). Estimation of variance of the general regression estimator: Higher level calibration approach. *Survey methodology*, *24*(1), 41-50.

Singh, S. (2003). *Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy* (Vol. 2). Springer Science & Business Media.

Sud, U. C., Chandra, H., & Gupta, V. K. (2014). Calibration approach-based regression-type estimator for inverse relationship between study and auxiliary variable. *Journal of Statistical theory and Practice*, *8*, 707-721.

Tracy, D. S., Singh, S., & Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology*, *29*(1), 99-104.