# GENERAL DISTANCE FORMULA ESTIMATION OF POPULATION TOTAL FOR UNEQUAL PROBABILITY SAMPLING DESIGNS WITH AUXILIARY VARIABLES

IBRAHIM ELABID IBRAHIM

UNIVERSITI TEKNOLOGI MALAYSIA

GENERAL DISTANCE FORMULA ESTIMATION OF POPULATION TOTAL FOR UNEQUAL PROBABILITY SAMPLING DESIGNS WITH AUXILIARY VARIABLES

IBRAHIM ELABID IBRAHIM

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

APRIL 2021

## DEDICATION

This study is devoted to those who have affected my life deeply, ,

*My Father, Mother, Brothers, and Sisters*

*My beloved wife Wafaa Elnasir Ibrahim*

*My cute daughters Raneem and Raheem*

*Ustaz Abd Alaziz Mohammed Ahmed Tangasawi*

# ACKNOWLEDGEMENT

First of all, I would like to sincerely thank my supervisor, Prof. Dr. Zuhaimy Bin Haji Ismail, during the long period of research, for all essential support, guidance, encouragement, advice, inspiration, friendship, and solving all academic and non-academic problems. This research will not be the same as that presented here without his continued support and interest.

Next, I would like to present a deep appreciation to the Sudan Ministry of Higher Education and Scientific Research, the University of Khartoum, and the Faculty of Mathematical Sciences for supporting my PhD thesis in Malaysia, and I also thank Universiti Teknologi Malaysia for its support.

I would like to thank my extended family, my father and mother with special thanks for their assistance, relentless encouragement, prayer and support in building my personality.

Finally, I would like to express my sincere gratitude to my wife, Wafaa, and to our daughters, Raneem and Raheem, for their patience, support and understanding during the lengthy study time. You are the source of my inspiration and I would never have been the person I am today without you.

# ABSTRACT

Sampling is a process or technique to obtain statistical information about a finite population by selecting a representative sample from that population, by using an appropriate sampling design. Furthermore, in the process, the required information about the units in the sample is measured and the inference about the unknown population parameters such as means, totals and proportions are done. This study is focused on estimating an unknown population total for one target variable using single or multiple auxiliary variables correlated with the target variable. This study also explores two classical estimators, namely the ratio estimator and the linear regression estimator, which are used as an alternative to the Horvitz Thompson estimator in the presence of a single auxiliary variable to estimate an unknown population total. The theoretical and empirical aspects were used to compare between these two estimators. The comparison was carried out based on the sample size and the correlation coefficient between the target variable and the auxiliary variable. The empirical study using the secondary data set for small and medium sample sizes shows that the linear regression estimator is more efficient compared to the ratio estimator when the correlation coefficient of the two variables is positive. For a large sample sizes, there are no significant differences between the two estimators. Also, the variance of both estimators decreases when the sample size increases. In contrast, if the correlation coefficient is negative, then any increase in the sample size leads to significant decrease in the variance estimate of the linear regression estimator. Meanwhile, for the ratio estimator, as the sample size is considerably increased, the variance of the estimator decreases. The simulation study showed that when the variable of interest has a strong negative correlation with the auxiliary variable irrespective of the sample size, the linear regression estimator provides an efficient estimate for the unknown population total relative to the ratio estimator. While, if the correlation coefficient between the variable of interest and the auxiliary variable is positive and within the range [0.75, 1], then the two estimators give a better estimate for the population total compared to the conventional estimators. However, the estimate of the total population obtained by the linear regression estimator is slightly more efficient than the ratio estimator. The most important idea in the estimation by using minimum distance measures is the quantification of the degree of closeness between the two data sets, such as sample data and the parametric distribution depends on an unknown parameter. A general distance formula is suggested in this research, based on the concept of the power divergence function, rather than that used by Deville and Särndal to measure the degree of closeness between the calibrated weights (new weights) and the classical design weights in Horvitz Thompson estimator. Derivation of the proposed general distance formula involved adding another constraint to the calibration equation constraints with respect to the sum of the classical sample design weights and the sum of sample calibrated weights. In order to generate a variety of distance measurements, the proposed formula was used to obtain a set of new weights that could be used to construct new estimators based on the inverse functions created by the proposed formula for estimating the total unknown population. Finally, the problems associated with calibrated weights produced by some distance measures, such as unrealistic or extreme weights are examined, leading to inaccurate estimates when these weights were handled instead of the design weights.

## ABSTRAK

Persampelan adalah proses atau teknik untuk mendapatkan maklumat statistik mengenai populasi terhingga dengan memilih sampel daripada populasi tersebut, dengan menggunakan reka bentuk persampelan yang sesuai. Seterusnya, dalam proses ini, maklumat yang diperlukan mengenai unit dalam sampel tersebut diukur dan inferens mengenai parameter populasi yang tidak diketahui seperti min, jumlah dan perkadaran dilakukan. Kajian ini difokuskan dalam menganggarkan jumlah seluruh populasi yang tidak diketahui menggunakan pembolehubah tunggal atau pembolehubah berganda bantu yang berkorelasi dengan pembolehubah sasaran. Kajian ini juga meneroka dua penganggar klasik, iaitu penganggar nisbah dan penganggar regresi linear, yang digunakan sebagai alternatif kepada penganggar Horvitz Thompson dengan kewujudan pembolehubah bantu tunggal untuk menganggarkan jumlah seluruh populasi yang tidak diketahui. Aspek teori dan empirikal telah digunakan untuk membandingkan antara kedua penganggar ini. Perbandingan dilakukan berdasarkan kepada ukuran sampel dan pekali korelasi antara pembolehubah sasaran dan pembolehubah bantu. Kajian empirikal menggunakan set data sekunder untuk ukuran sampel kecil dan sederhana menunjukkan bahawa penganggar regresi linear lebih berkesan berbanding penganggar nisbah apabila pekali korelasi kedua-dua pembolehubah itu adalah positif. Bagi saiz sampel yang besar, tidak terdapat perbezaan ketara antara kedua-dua penganggar. Juga, varians bagi kedua-dua penganggar menyusut apabila saiz sampel meningkat. Sebaliknya, jika pekali korelasi adalah negatif, maka setiap peningkatan dalam saiz sampel menyebabkan penyusutan yang ketara dalam anggaran varians penganggar regresi linear. Manakala, bagi penganggar nisbah, kerana saiz sampel jauh meningkat, varians penganggar menyusut. Kajian simulasi menunjukkan bahawa apabila pembolehubah yang diminati mempunyai korelasi negatif yang kuat dengan pembolehubah bantu tanpa mengira ukuran sampel, penganggar regresi linear memberikan anggaran yang berkesan untuk jumlah seluruh populasi yang tidak diketahui berbanding dengan penganggar nisbah. Sementara itu, jika pekali korelasi antara pembolehubah yang diminati dan pembolehubah bantu positif dan berada dalam julat [0.75, 1], maka kedua-dua penganggar memberikan anggaran yang lebih baik untuk jumlah seluruh populasi berbanding dengan penganggar lazim. Walau bagaimanapun, anggaran jumlah seluruh populasi yang diperoleh oleh penganggar regresi linear sedikit lebih cekap daripada penganggar nisbah. Idea yang paling utama dalam perkiraan dengan menggunakan ukuran jarak minimum adalah pengukuran tahap kedekatan antara dua set data, seperti data sampel dan taburan parametrik bergantung pada parameter yang tidak diketahui. Rumus jarak umum telah dicadangkan dalam penyelidikan ini, berdasarkan kepada konsep fungsi perbezaan daya, berbeza daripada yang digunakan oleh Deville dan Särndal untuk mengukur tahap kedekatan antara pemberat yang ditentukur (pemberat baru) dan pemberat reka bentuk klasik penganggar Horvitz Thompson. Penerbitan rumus jarak umum yang dicadangkan melibatkan penambahan kekangan lain pada kekangan persamaan penentukuran berkenaan dengan jumlah pemberat reka bentuk sampel klasik dan jumlah sampel pemberat yang ditentukur. Bagi menghasilkan pelbagai ukuran jarak, rumus yang dicadangkan telah digunakan untuk mendapatkan satu set pemberat baharu yang boleh digunakan untuk membangun penganggar baharu berdasarkan kepada fungsi songsang yang dijana oleh rumus yang dicadangkan untuk menganggarkan jumlah seluruh populasi yang tidak diketahui. Akhirnya, masalah yang berkaitan dengan pemberat yang ditentukur terhasil daripada beberapa ukuran jarak, seperti pemberat tidak realistik atau ekstrem diteliti, yang menyebabkan anggaran kurang tepat ketika pemberat ini dikendalikan dan bukannya daripada pengunaan pemberat reka bentuk.

# TABLE OF CONTENTS

| TITLE | PAGE |
|---|---|
| **DECLARATION** | **iii** |
| **DEDICATION** | **iv** |
| **ACKNOWLEDGEMENT** | **v** |
| **ABSTRACT** | **vi** |
| **ABSTRAK** | **vii** |
| **TABLE OF CONTENTS** | **viii** |
| **LIST OF TABLES** | **xii** |
| **LIST OF FIGURES** | **xiii** |
| **LIST OF ABBREVIATIONS** | **xiv** |
| **LIST OF SYMBOLS** | **xv** |
| **LIST OF APPENDICES** | **xvii** |
| **CHAPTER 1 INTRODUCTION** | **1** |
| 1.1 Overview | 1 |
| 1.2 Problem Background | 2 |
| 1.2.1 Target and Auxiliary Variables | 4 |
| 1.2.2 Statistical Minimum Distance Measures | 8 |
| 1.3 Problem Statement | 9 |
| 1.4 Research Objectives | 12 |
| 1.5 Motivation for Research | 13 |
| 1.6 Significance and Contribution of the Study | 14 |
| 1.7 Scope of the Research | 15 |
| 1.7.1 Scope of the Sampling Strategy | 15 |
| 1.7.2 Scope of the Data | 15 |
| 1.7.3 Scope of the Methods | 16 |
| 1.8 Limitations of the Study | 17 |
| 1.9 Organization of the Thesis | 17 |

**CHAPTER 2 LITERATURE REVIEW 19**

2.1 Introduction 19

2.2 Overview of Unequal Probability Sampling Designs 19

2.3 Sampling with Probability Proportional to Size 21

2.4. Estimation of Population Total Using HT Estimator 25

2.5. Estimation of the Population Total under $\pi \boldsymbol{PS}$ Scheme 27

2.6. Inference under Three Different Approaches 28

2.6.1 Design-Based Approach 29

2.6.2 Model-Based Approach 29

2.6.3 Model-Assisted Approach 31

2.7 Calibration Estimation for the Population Total 31

2.8. Summary 44

**CHAPTER 3 RESEARCH METHODOLOGY 45**

3.1 Introduction 45

3.2 Research Design 46

3.3 Research Framework 46

3.3.1 Phase 1: Research Formulation 48

3.3.2 Phase 2: Estimation Methods 49

3.3.3 Phase 3: Validation and Testing 50

3.3.4 Phase 4: Conclusion and Report Writing 50

3.4 The Concept of Sampling Design 52

3.4.1 Sample Selection Schemes 52

3.4.2 Sampling Design and Inclusion Probabilities 57

3.5 Estimation of Population Total under Unequal Probability Sampling 60

3.5.1 Horvitz Thompson Estimator (Classical Scenario of $\pi$PS) 61

3.5.2 The Properties of the Inclusion Probabilities 68

3.6 The Proposed Estimator Based on New Weights 71

3.7 Summary 72

**CHAPTER 4 COMPARISON OF TWO CLASSICAL ESTIMATORS: RATIO AND LINEAR REGRESSION** **75**

4.1 Introduction 75

4.2 Ratio Estimation 76

4.3 Linear Regression Estimator 80

4.4 The Description of Data 82

4.5 Comparison between Two Estimators 86

4.6 Simulation Study 89

4.7 Summary 93

**CHAPTER 5 CALIBRATION ESTIMATION BY USING POWER DIVERGENCE FUNCTION** **95**

5.1 Introduction 95

5.2 Calibration Estimator Based on Auxiliary Variables 96

5.3 Determining the Calibration Weights 99

5.4 The General Distance Measure 104

5.5 Quantification of Distance 108

Definition 5.1 109

Definition 5.2 109

5.5.1 Disparity Function Properties 110

5.6 Derivation of the Inverse Functions $F_k(u), F(q_k u)$ using Power Divergence Function 111

5.6.1 The Power Divergences Function 111

5.6.2 Derivation of General Formula of $F(q_k u)$ From the PD Function 112

5.7 Validation of the Efficiency of the PGFDM 114

5.8 Treatment of Inappropriate Calibrated Weights Produced by PGFDM 115

5.9 Summary 119

**CHAPTER 6 CONCLUSION AND RECOMMENDATIONS** **121**

6.1 Summary of the Research 121

6.1.1 Ratio and Linear Regression Estimator 122

6.1.2 General Formula for Distance Measure 123

6.2 Contribution of the Study 124

6.3 Future Work 124

**REFERENCES** **127**

**LIST OF PUBLICATIONS** **163**

## LIST OF TABLES


| TABLE NO. | TITLE | PAGE |
|---|---|---|
| **Table 1-1** | The Key Variations between the Target Variables and the Auxiliary Variables | 5 |
| **Table 1-2** | Illustration of the target variables and the auxiliary variables for six hypothetical studies | 6 |
| **Table 2-1** | The Main Differences between the Design Based, the Model Based and the Model Assisted Approaches | 37 |
| **Table 2-2** | The Previous Works Related to the Calibration Estimator | 39 |
| **Table 3-1** | Four different presentation of a sample with effective sample size 1 | 55 |
| **Table 3-2** | Four different presentation of a sample with effective sample size 2 | 55 |
| **Table 3-3** | Four different presentation of a sample with effective sample size 3 | 56 |
| **Table 3-4** | The weekly wages $X$ and expenditure $Y$ in five households | 65 |
| **Table 3-5** | Selected Sample and Probability of Each Selection | 67 |
| **Table 4-1** | Simple random sample of size 30 generated by Algorithm 4.1 | 83 |
| **Table 4-2** | Simple random sample of size 50 generated by Algorithm 4.1 | 84 |
| **Table 4-3** | Estimate of the total amount of pollution produced by nitrogen oxides using $\ln NO_x Em$ as an auxiliary variable | 87 |
| **Table 4-4** | Estimate of the total amount of pollution produced by nitrogen oxides using square root of wind speed (m/s) as an auxiliary variable | 87 |
| **Table 4-5** | Population total estimate when the correlation coefficient between the variable of interest and the auxiliary variable is within the range [0.5, 0.75] | 91 |
| **Table 4-6** | Population total estimate when the correlation coefficient between the variable of interest and the auxiliary variable is within the range [-0.75, - 0.5] | 91 |
| **Table 4-7** | Population total estimate when the correlation coefficient between the variable of interest and the auxiliary variable is within the range (0.75, 1] | 91 |
| **Table 4-8** | Population total estimate when the correlation coefficient between the variable of interest and the auxiliary variable is within the range [- 1, - 0.75) | 92 |
| **Table 5-1** | Distance measures presented by Deville and Särndal (1992) | 107 |

## LIST OF FIGURES


| FIGURE NO. | TITLE | PAGE |
|---|---|---|
| **Figure 1.1** | Classification of Sampling Designs | 7 |
| **Figure 3.1** | Research Framework Phases | 47 |
| **Figure 3.2** | Research Framework | 51 |
| **Figure 3.3** | Illustrate Designs A | 59 |
| **Figure 3.4** | Illustrate Designs B | 59 |

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| SRS | - | Simple Random Sampling |
| *PPSWR* | - | Probability Proportional to Size With Replacement |
| *PPSWOR* | - | Probability Proportional to Size Without Replacement |
| HT | - | Horvitz Thompson |
| HH | - | Hansen Horvitz |
| GREG | - | Generalized Regression Estimator |
| JB | - | Johor Bahru |
| WR | - | Sampling With Replacement |
| WOR | - | Sampling Without Replacement |
| LR | - | Linear Regression |
| RE | - | The relative efficiency between two estimators |
| *iid* | | Independent identically distributed |
| PGFDM | | The proposed general formula of distance measure |
| *UPSWOR* | | Unequal probability sampling without replacement |
| BS | | Bernoulli sampling |
| PS | | Poisson sampling |
| OS | | Ordered sample |
| RSS | | Rank Set Sampling |
| CV | | Coefficient of variation |
| R | | The ratio between the two variables |
| C | | Disparity generating function |
| PDF | | Power divergence function |

## LIST OF SYMBOLS

| | |
|---|---|
| $P$ | Sampling Design |
| $\hat{T}$ | Estimator of an Unknown Population Parameter $T$ |
| $n_s$ | Random sample size |
| $n$ | The Fixed Sample Size |
| $s$ | The sample $s$ |
| $v(s)$ | The effective samples size |
| $p(s)$ | The probability of selecting the sample $s$ |
| $\Omega$ | All possible of random samples that can be selected from a population |
| $I_i$ | The Inclusion Indicator for the Unit $i$ |
| $I_{ij}$ | The Inclusion Indicator for the Units $i, j$ |
| $\underline{I}$ | The Design Vector |
| $M(n; p_1, p_2, ..., p_N)$ - | Multinomial Distribution |
| $\pi$PS | Probability proportional to size without replacement under Unequal Probability |
| $\pi_k$ | The First Order inclusion Probability for the $k^{th}$ unit |
| $\pi_{ij}$ | The Second Order Inclusion Probability |
| $d_k$ | The traditional sampling weight for the $k^{th}$ unit |
| $\hat{Y}_{HH}$ | Hansen Horvitz Estimator |
| $\hat{Y}_{HT}$ | Horvitz Thompson Estimator |
| $\hat{t}_{Das}$ | Das Estimator for Population Total |
| $w_k$ | The calibrated weight for the $k^{th}$ unit |
| $q_k$ | The Individual Weight |
| $C_{pu}$ | The class of $p-$unbiased estimators |
| $C_{\zeta u}$ | The class of $\zeta-$unbiased estimators |
| $\rho$ | The correlation coefficient between two variables |

| | |
|---|---|
| $\hat{Y}_R$ | Ratio estimator |
| $\hat{Y}_{LR}$ | Linear regression estimator |
| $RE(\hat{T}_1, \hat{T}_2)$ | Relative efficiency of $\hat{T}_1$ with respect to the $\hat{T}_2$ |
| $\boldsymbol{\lambda}$ | The vector of Lagrange multipliers |
| $A_{N\times p}$ | The data matrix of $N$ observations with $p$ auxiliary variables |
| $\hat{\boldsymbol{B}}$ | The vector of weighted estimator of multiple regression coefficients |
| $D(w_k, d_k)$ | The distance measure between the calibrated weights and the classical weights |
| $\delta(x)$ | Pearson residual at $x$ |
| $F_k(u)$ | The inverse function |
| $F(q_k u)$ | The simplest form of the inverse function |
| $\rho C(d_n, f_\theta)$ | Disparity between the vectors $d_n$ *and* $f_\theta$ |
| $\hat{Y}_c$ | Calibration estimator |

## LIST OF APPENDICES


| APPENDIX | TITLE | PAGE |
|---|---|---|
| Appendix A | Simple random sample of size 100 generated by Algorithm 4.1 | 135 |
| Appendix B | Simple random sample of size 200 generated by Algorithm 4.1 | 139 |
| Appendix C | Simple random sample of size 400 generated by Algorithm 4.1 | 146 |
| Appendix D | R Programming Code for Empirical Study | 159 |
| Appendix E | R Programming Code for Simulation Study | 160 |

# CHAPTER 1

# INTRODUCTION

## 1.1 Overview

This research is a hybrid of theoretical and practical concepts. A thorough background in probability theory and mathematical statistics is needed to understand the basic concepts of sampling designs and estimation of population parameters. Sampling survey is a process used to describe or summarize the characteristics of a particular population or making some inferences about the population parameters based on data collected from the sample. The essence issue of the sampling survey is how to select a sample from a particular population, and how to estimate the unknown parameters for this population.

A survey can be defined as any activity performed by using well-defined concepts, methods and procedures to collect information from a whole population or a part of the population. Surveys are usually conducted by governments, organizations, businesses, and researchers in order to describe or summarize a population or to make some inferences from the variables under study based on the sample selected (Franklin and Walker, 2003). There are two types of surveys: complete (census) and sampling surveys. The complete survey (census), in which the information is collected from all units of the population. For sample survey, the sample units are collected from a portion of the target population called a sample. A complete survey provides the actual values of the population parameters of the study variables. In a sample survey, the true values of population parameters are difficult to obtain. The main purpose of using the sample survey rather than the complete survey (census) is to reduce the cost of study, save time, effort and increased the accuracy of estimates (Cochran, 1977, Singh, 2003).

An important term in the sampling survey field is a sampling strategy, defined as a combination of a sampling design $(P)$ and an estimator of population parameter

$(\hat{T})$. A sampling design is used to select a sample units, while an estimator is used to estimate an unknown population parameter based on the sample was selected according to the sampling design. In other words, a sampling strategy is a process of selecting a sample, coupled with a procedure to estimate unknown population parameters based on this sample. Sampling strategy involves two stages, design and estimation stages.

The design stage determines the design used to draw the elements from the entire population. The estimation stage describes how to estimate the parameters of a particular population, especially a total. In this study, in the presence of a single auxiliary variable or several auxiliary variables, estimation of population total for a single target variable is carried out under unequal probability sampling designs.

A review of the basic concepts and theorems linked to the sampling designs is needed to develop an effective sampling strategy, which are given in Chapter 3 in detail. This study focuses mainly on the estimation stage, in particular on two traditional estimators based on a single auxiliary variable, namely the ratio estimator and linear regression estimator. In addition, the calibration estimation method in presence of several auxiliary variables would also be studies through modifying the traditional weights (design weights) by calibrated weights (new weights).

## 1.2 Problem Background

Sampling can be defined as a technique or method for selecting a representative sample from a specific population, with the goal of estimating the unknown parameters of the population characteristics such as mean, total and proportion. A sample is said to be a representative sample or a random sample when the units are selected randomly. Generally, a sample is selected based on a specific probability mechanism referred to as sampling design. Whether the sample is selected for a descriptive or analytical purpose, an appropriate sampling design should be used to obtain results with high precision. Sampling designs fall into two categories, probability sampling and non-probability sampling designs.

Probability sampling designs are applied when every unit in the population has a known chance for being selected in a sample. Thus, using these designs for selecting a sample guarantee no part of the population is ignored before the sample selection process. Simple random sampling (SRS), systematic, stratified and cluster sampling designs are types of probability sampling design. In non-probability sampling, the selection of units is subjective (not random) and totally depends on the researcher or data collector. Quota sample, judgmental sample, and convenience sample are examples of non-probability sampling schemes. Although, these sampling techniques provide fast, easy, and low cost ways of selecting samples, but give less accurate estimates and are less used (Cochran, 1977, Singh, 2003, Arnab, 2017). Probability sampling designs can be classified into:

**Equal probability sampling designs**: These designs give equal probabilities for the all population units to select in any sample. In other words, all possible samples have an equal chance of being selected as a sample. The simplest example for these designs is SRS. The key problem associated with these schemes is that the essential differences in the values of population units are not taken into account, so selecting a sample according to these designs produce results which are not consistent, particularly when the units have significant variation with respect to their values (Sampath, 2001, Särndal et al., 2003).

**Unequal probability sampling designs**. These designs are applicable when the population units do not have equal probabilities for being selected in the sample. This research is performed according to sampling designs with unequal probability. These designs can provide better estimates compared to sampling schemes with equal probability.

Auxiliary information may be used in a design stage or estimation stage or both in unequal probability sampling designs. Using this information, an improvement in the efficiency of estimating population parameters can be achieved through a suitable sampling design. The next Section is dedicated to highlight on auxiliary variables.

### 1.2.1 Target and Auxiliary Variables

The simple use of auxiliary information is one of the main contributions for the paper presented by Neyman in 1934. Before selecting a sample, Neyman suggested dividing the population into strata, then selecting a probability sample from each stratum. The official use of auxiliary information in the sampling survey traces back to the early decades of the last century when the foundations of the sampling theory were laid down. (Holmberg, 2003).

Cochran (1942) presented an early reference which introduces auxiliary information that can be used in the estimation procedures. Also he proposed the regression method of estimation in detail.

Target variables (variables of interest) are that variables we wish to make some statistical inference around them in a particular study. Whereas the auxiliary variables are the variables that are directly or indirectly related to the target variables, used to improve the precision of the estimation of unknown population parameters, such as population total, or to predict the values of the target variables. An auxiliary variable is usually any variable for which information is available prior to the information. The values of the auxiliary variables for the all population units must be known before selecting any random sample from the population. Using of suitable auxiliary variables in the estimation stage results in considerable reduction in the variance estimate of the unknown population parameters.

In practice, knowing only the values of the sample units and the population totals of the auxiliary variables is sufficient for making any statistical inference about unknown population total. Usually, referred to the target variable by $Y$ while the auxiliary variable by $X, Z, \ldots$ or $X_1, X_2, \ldots, X_p$. The primary objective of using auxiliary variables is to obtain accurate estimates leading to many sampling and estimation methods, such as optimum allocation in stratified sampling designs, probability proportional to size, and regression or calibration estimators, all of which rely on properties of these variables.

Auxiliary variables are used at the design stage or at the estimation stage or in both stages in a wide variety of ways. Many works have been published to illustrate this wide variety of using auxiliary variables, such as such as, Laaksonen (2006) proposed a particular scheme of auxiliary data services to systematize auxiliary variables according to different types. Estevao and Särndal (2002) showed that there are ten different cases where auxiliary information may be used for calibration in two-phase sampling.

According to (Franklin and Walker, 2003), before selecting a sample, if auxiliary information is available and easily accessible from a reliable sources, in this case it can be used in estimation stage for two main reasons: The first reason is that the survey results always need to be as close as possible to the actual population parameters or parameters estimates from another, more accurate survey. The second reason is to increase the efficiency of the parameters estimates

There are several aspects that can be used to differentiate between the target variables and the auxiliary variables, such as cost, effort, source of availability, etc. Table 1.1 shows the key differences between the variables of interest and the auxiliary variables based on some aspects that are related to a particular study.

**Table 1-1** The Key Variations between the Target Variables and the Auxiliary Variables

| Study aspects | Auxiliary variables | Target variable |
|---|---|---|
| Cost of data collection | Less | High |
| Effort of data collection | Less | More |
| Sources of availability | Previous studies, secondary data, current or past survey | Experiments, current survey |
| Interest | Less | More (main) |
| Error in measurements | Less | More |
| Sources of error | Fewer | More |

In order to explain the nature of the target variables and auxiliary variables, Table 1.2 shows 6 hypothetical studies, with the determination of the target variable and auxiliary variable for each study.

**Table 1-2** Illustration of the target variables and the auxiliary variables for six hypothetical studies

| The study | Target variable | Auxiliary variable |
|---|---|---|
| Agricultural survey | The annual productivity of a specific crop. | The area cultivated, soil fertility |
| Family survey | The total amount spent by household on food consumer | Household income, size of family, and type of food. |
| Health survey | Measures of risk factors, health status | Tobacco use, alcohol use, diet, and physical exercise |
| Environmental survey | Total amount of pollution produced from factories in a particular area. | Factory size, number of worker, number of units produced last year |
| Business survey | Income, profit, number of employees | Annual sales, number of employees and number of locations |
| Market survey | Satisfaction levels for a particular product. | Age, gender, education level |

There are three basic requirements at the estimation stage for the successful utilization of auxiliary variables:

i. The auxiliary variables are well correlated with the target variable.

ii. The sources of information concerning the population must be accurate and easy to collect.

iii. The auxiliary information must be obtained from all units selected in the sample, and assumed the population totals of the auxiliary variables are known.

Improvements in the estimation of population parameters using auxiliary variables rely on the degree of correlation between the target variable and the auxiliary variables available. Although the issue of using auxiliary data in the sampling survey began to attract interest in the 1930s last century, it is still an active topic of research.

An important concept associated with a sampling design is a sampling scheme. The sampling scheme is defined as a set of rules that are used to give a detailed description of how to include the population elements in the sample so that the probabilities of sampling design are fulfilled.

A classification of the all sampling designs that are mentioned in the previous section are shown in the Figure 1.1

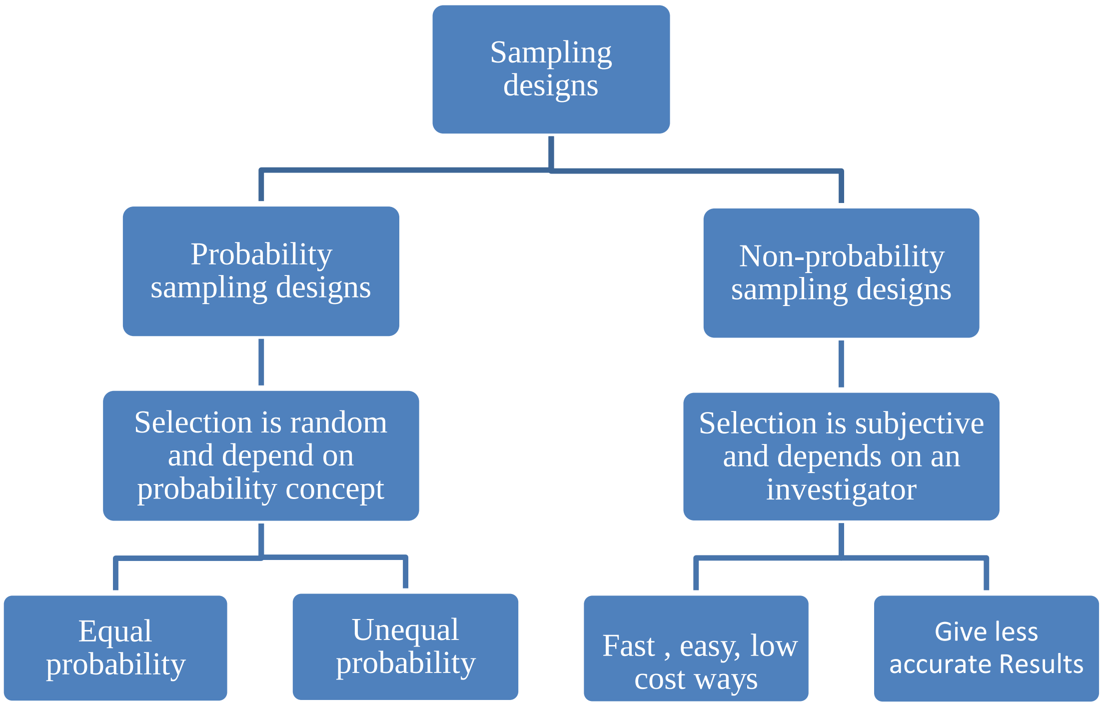


**Figure 1.1** Classification of Sampling Designs

### 1.2.2 Statistical Minimum Distance Measures

Estimating unknown population parameters, especially the total population, in many cases involves using new weights instead of those used in traditional estimators such as the Horvitz Thompson (HT) estimator. Minimum distance measures are one of the most effective statistical tools for obtaining these weights. Measuring the degree of closeness between two separate data sets, such as sample data and the parametric model used to suit this data, is the key concept of statistical distance measures between two data sets. The degree of closeness between the data and the model is described by some of statistical distance measures. Estimation by using the minimum distance measures was introduced by Wolfowitz in the middle of the last century (1952, 1953, 1954, 1957). He studied minimum distance estimators as class of estimators, looked at the case when the sample size is large, demonstrated that under general conditions such estimators are highly reliable, also used minimum distance measures in goodness-of-fit tests. The minimum distance was outlined by William (1981) as a general estimation method with desired properties. An extensive bibliography, categorised by subject matter, was also provided. Vajda (1989) and Pardo (2018) provided useful treatments based on divergence measures for statistical inferences. In statistical inference, statistical distance measures can be used across two types of distances:

i. Distances between two distribution functions, such as distance between the empirical distribution function and the model distribution function.

ii. Distances between two probability density functions, such as distance between some nonparametric density estimates obtained from a data and the model probability density function.

An important relevant term related to statistical distance measures is the quantification of the amount of discrepancy between the data and the model used to fit this data through a suitable divergence measure. Quantification of the amount of discrepancy is one of the key components of statistical modelling.

The idea of statistical distance was used in sampling survey, especially for estimating an unknown population total in the presence of auxiliary variables.

Calibration estimation was presented by Deville and Särndal (1992), by creating a suitable distance measure between the design weights that are contained in HT estimator and the calibrated weights. Calibration estimator is obtained by minimizing the suitable distance measure between these two sets of weights.

This study focuses mainly on the issue involved in estimating an unknown population total of a single target variable under unequal probability sampling designs in the presence of auxiliary variables. Two methods of estimation were employed in this study. The classical scenarios of estimation (which includes linear regression estimator, ratio estimator and Horvitz-Thompson Estimator) were used in first category of estimation. The estimation by these scenarios are based on a single auxiliary variable. The second category is estimation by calibration method which based on vector of auxiliary variables using different manner of statistical distance measure rather than that was used in the previous studies.

## 1.3 Problem Statement

In many cases, equal probability sampling schemes, such as simple random sampling (SRS) and systematic random sampling, are not suitable for selecting a random sample because it is affected by the problems related to the population units, particularly when the values of the population units have essential differences. For this reason, equal probability sampling schemes do not seem to be an appropriate procedure for selecting a representative sample. On the contrary, selection of units using unequal probabilities could improve the efficiency of estimating the population parameters through reducing the variance estimate, in many circumstances.

Estimation of population parameters can be improved by using unequal probability sampling designs rather than equal probability sampling designs. The latter designs take into account the essential differences between the population units before selecting any sample (Sampath, 2001). Each unit of the population has an inclusion probability in any sample. The inclusion probabilities determined by using an appropriate sampling design, such as Poisson sampling design, Conditional Poisson sampling, Pareto sampling, etc.

Most of the previous works in the sampling design begun in the early 1940s were devoted to the estimation of population parameters under two sampling schemes, especially the population total, taking into account a single auxiliary variable. One of the scheme, is the probability proportional to size with replacement (*PPSWR*) and the other is the probability proportional to size without replacement (*PPSWOR*). The two well-known estimators for estimating an unknown population total according to *PPSWR* and *PPSWOR* schemes are: Hansen Horvitz (HH) estimator proposed by Hansen and Hurwitz (1943) and Horvitz Thompson (HT) estimator suggested by Horvitz and Thompson (1952) respectively. In order to improve these two estimators, a number of sample selection schemes and estimators were later proposed. Madow (1949), Narain (1951), Raj (1956 a), Roy and Chakravarti (1960), Murthy (1957), Srivenkataramana and Tracy (1980) , to name but a few.

Many studies have been conducted to address the problem related to difficulties in computing the variance estimate of the HT estimator using the second order inclusion probabilities, such as the generalized regression estimator (GREG) suggested by Cassel et al. (1977), calibration estimators given by Deville and Särndal (1992), calibration estimation by empirical likelihood method proposed by Chen and Qin (1993) and extended by Kim (2009). Singh et al. (1999) suggested new calibration approach in order to improve the variance estimate. Guggemos and Tillé (2010) suggested a class of model-assisted estimators based on the penalization calibration. The underlying idea in this method is to reduce the calibration constraints by releasing a few of them and replacing with a penalty term. The penalization added minimization to the distance measure. They introduced the concept of penalized calibration as a combination of traditional calibration and adjusted calibration. This method provides a flexible estimation technique giving improved estimates, when the auxiliary information, particularly, is not fully suitable or overly abundant to be completely used. Singh and Sedory (2016) gave the two-steps technique to calibrate the design weights. In stage one, the calibration weights are made proportional to the design weights based on a particular sample. The research area in the estimation of population parameters in sampling survey, in the presence of several auxiliary variables by using the calibration estimation technique is fertile and need further work

In all previous works that are mentioned above, most attention was paid to estimation of population total in the presence of several auxiliary variables based on the calibration estimation proposed by Deville and Särndal (1992). In other words, all these contributions were done by using the same distance measures between design weighs and calibrated weights used by Deville and Särndal (1992). In contrast, the literature of estimation of population parameters suffers from acute lack of studies related to the issue of statistical distance measures that can be used to find the calibrated weights, this is as a results of using the same distance measures in the all previous works. No study was devoted to the subject of statistical measurements of distances between the design weights and calibrated weights, except that Deville and Särndal used some distances measures. Therefore, filling this gap is one of our main objectives in this research. The study focuses specifically on the issue of statistical distances measures that can be used to determine the amount of difference between design weights and calibrated weights.

In this study, conventional estimation methods are revisited based on the availability of a single auxiliary variable. In order to estimate an unknown population total, two conventional estimators, namely the ratio and regression estimators, are compared theoretically and practically by using empirical data and simulation technique. In addition, proposed a general formula of statistical distance measure, used to measure the distance between the new weights and traditional weights. New weights can be obtained via the general formula and then the new weights can be used to create new estimators for estimating an unknown population total. The key objective sought of this study is to obtain a variance estimates that are free from the obstacles relevant to the inclusion probabilities of the second order. As in variance estimates for the classical scenario of unequal probability sampling designs, calculation of variance estimates are tedious process and computationally cumbersome. The study of the problematic part of variance estimates and issues related to statistical distance measures lead to some important questions related to problem statement represented in the research questions which include:

**QR 1:** How much impact does the sample size and correlation coefficient have on the estimate of population total to be computed using the ratio and regression estimators?

**QR 2:** Which general distance measure is required to find the calibrated weights under some constrains related to the auxiliary variables?

**QR 3:** How can the special distance measures be generated from the general formula?

**QR 4:** How can the negative and unrealistic weights for some distance measure be treated?

## 1.4 Research Objectives

One of the main goals of this study is to re-examine the conventional estimation methods using two well-known classical estimators, linear regression and ratio estimators, used for estimating the total population in the presence of a single auxiliary variable. The impact of sample size and correlation coefficient between the target variable and auxiliary variable on the two conventional estimators would be examined, then the results would be compared from a theoretical and practical point of view. Furthermore, if more than one auxiliary variable is available from a reliable sources, calibration estimation is used by incorporating auxiliary variables into the estimation process. To measure the degree of the disparity between calibrated weights (new weights) and design weights (traditional weights), a general distance formula is proposed. In order to deal with the problems related to the estimation of an unknown population total mentioned in the problem statement, we need to define the research objectives. The main objectives of this study can be determined as follows:

i. To investigate and analyse the effect of sample size and correlation coefficient on the efficiency of the ratio estimator and linear regression estimator based on a single auxiliary variable.

ii. To design a general distance formula that can be used to obtain the calibration weights for some well-known distance measures.

iii. To construct a new estimator based on the new weights generated from the general distance formula.

iv. To treat the unrealistic values of new weights produced by some special distance measures.

## 1.5 Motivation for Research

Both academic and personal motivations inspired the candidate for performing this study. First of all, sampling survey is a significant tool for scientific research, and also it counts as a corner stone for many scientific researches in various fields, such as medicine, economics, management, marketing, forestry, education, industry, pharmacy, etc. As a consequence, any progress in this area of study, such as development of sampling designs or population parameter estimation methods, may have a huge impact on the precision of the decisions made in relation to a number of practical problems.

Secondly, making any statistical decision depends on the data gathered using a specific sampling design. Using an unsuitable sampling design lead to wrong decisions being made. In addition, the precision of estimate is very important issue for all tacticians and researchers. This motivates the candidate to further improve the methods of estimation of population parameters by reducing the possibility of using an unsuitable sampling design.

Third, Estimation by an appropriate statistical minimum distance method is one of the most efficient methods in estimating the unknown population parameters, particularly when auxiliary variables are available and associated with the variable of interest

Finally, in many situations, we need to get answers to certain specific questions, which cannot be obtained merely through laboratory experiments or simply using economic, mathematical, or statistical formulation. The answers can be conveniently obtained by sampling survey.

## 1.6 Significance and Contribution of the Study

In equal probability sampling designs, the sampling design used to select a sample and estimation procedures are mainly depend on the study variable only. In fact, there are many cases where additional information from some variables related to the study variable is available. Additional information from current or past surveys or secondary sources can be made available from books, magazines, journals, etc. Such additional information is called auxiliary variables. These variables play a prominent role in improving estimation accuracy. This study looks at the topic of estimating an unknown population total based on two different scenarios. In the first scenario where there is a single auxiliary variable, two classical estimators (linear regression estimator and ratio estimator) are used to increase the precision of the estimate of unknown population total. Whereas, estimation of unknown population total is treated by the calibration estimation in the second scenario when several auxiliary variables are available. The expected contributions from this study are as follows::

i. Description of the effect of the sample size and correlation coefficient on the efficiency of population total estimate by ratio estimator and linear regression estimator.

ii. Provide a theoretical overview on the calibration estimation based on the minimum distance approach.

iii. Presenting a new general distance measure used to generate several well-known distance measures.

iv. Connecting between classical estimators and calibration estimator for the ratio estimator and regression estimator.

## 1.7 Scope of the Research

There are three key aspects to the scope of this study, including the scope of the sampling strategy, the scope of the data, and the scope of the methods. These aspects cover the sampling strategy's theoretical context, the data used in the study, and the methods used to estimate an unknown population total for a particular target variable. The following subsections identify the three aspects of the study scope.

### 1.7.1 Scope of the Sampling Strategy

The scopes of sampling strategy in this research are as follows:

i. This study is restricted to methodology of unequal probability sampling designs.

ii. Auxiliary information is used in design and estimation stages.

iii. Study mainly focuses on the estimation stage.

### 1.7.2 Scope of the Data

The data used in this study are hypothetical data and secondary data, described as follows:

i. Hypothetical data is used in some numerical examples in order to promote the comprehension of some fundamentals relevant to sampling strategy.

ii. A secondary data is used to compare two traditional estimators; the ratio estimator and linear regression estimator.

iii. Simulated data based on one variable of interest and two auxiliary variables, one having a positive correlation with the variable of interest and the second with a negative correlation to compare the efficiency of the estimators.

### 1.7.3 Scope of the Methods

This study is focused on the estimation of unknown population total in the presence of a single auxiliary variable or several auxiliary variables. For the treatment of both cases, two scenarios of estimation are established. The first scenario is based on a single auxiliary variable and the second is on several auxiliary variables. In this study many methods are used, which are as follows:

i. Five random samples of sizes 30, 50, 100, 200, and 400 would be chosen from a specific secondary data and simulation study to compare between the ratio estimator and the linear regression.

ii. Comparison between these estimators also is conducted by using simulation technique.

iii. Disparity measures are used to investigate the closeness between calibrated weights and designs weights.

iv. Calibrated weights are obtained using the optimization method, via minimizing the distance measure between calibrated weights and design weights.

iv. The proposed general distance formula is based primarily on a theoretical approach.

## 1.8 Limitations of the Study

This study has some statistical limitations which may influence the research findings. The limitations are briefly stated below:

i. Estimation of population total is carried out for only one target variable based on the auxiliary information from a single auxiliary variable or several auxiliary variables.

ii. The focus of the study is only on the auxiliary variables associated with the variable of interest, and other variables are ignored.

iii. Inference is limited to the design-based approach only.

## 1.9 Organization of the Thesis

This section describes how the thesis is arranged according to chapters. The thesis includes six chapters listed below:

**Chapter 1**, **Introduction:** This chapter gives a general introduction to the topic of the proposed research work. It covers brief overview of issues related to the research, problem background, problem statement, research objectives, motivation for research, significance of the study, scope of research, and the limitation of the study.

**Chapter 2, Literature Review:** This chapter presents a literature review of estimation of population total under unequal probability sampling designs. It covers an overview of unequal probability sampling designs, sampling with probability proportional to size, estimation of population total using HT estimator, estimation of population total based on probability proportional to size without replacement under unequal probability sampling designs, inference under three different approaches, and calibration estimation. Furthermore, this chapter also reviews the previous works related to unequal probabilities sampling designs and estimation of population total under unequal probability sampling designs, based on single auxiliary variable or multivariate auxiliary variables.

**Chapter 3, Research Methodology:** This chapter mainly presents the direction of the study and overview of the methods used. It includes the research framework, the basic concepts of sampling designs, estimation of population total based on unequal probability sampling designs, and finally, a new estimator based on new weights is proposed.

**Chapter 4, Comparison of Two Classical Estimators:** To address the first objective of research, in this chapter an empirical study and simulation study are carried out to estimate population total on two classical scenarios. Two traditional estimators, ratio and linear regression estimator, are revisited and used to estimate population total parameter, when a single auxiliary variable is available. The calculations obtained from the two estimators are presented and compared from a theoretical and empirical standpoint.

**Chapter 5, Calibration Estimation Using Power Divergence Function**: This chapter covers in detail, the last three objectives of the research. The calibration estimation of population total is examined. A general distance measure formula is proposed to measure the distance between classical and calibration weights. The proposed formula is then used to generate other notable distance functions. The new weights are then used to construct a new estimator, in some cases where the new weights are negative or unrealistic.

**Chapter 6**, **Conclusion and Recommendations:** This chapter provides the conclusions of the research work discussed throughout of this study, the chapter also presents and highlights the contributions of the research and put forward recommendations for future work.

## REFERENCES


Al-Jararha, J. & Ebrahem, M. a.-H. (2012). A Ratio Estimator under General Sampling Design. *Austrian Journal of Statistics,* 41**,** 105–115.

Ali, S. M. & Silvey, S. D. (1966). A General Class of Coefficients of Divergence of One Distribution from Another. *Journal of the Royal Statistical Society: Series B (Methodological),* 28**,** 131-142.

Alka, Rai, P. K. & Qasim, M. (2019). Two-Step Calibration of Design Weights under Two Auxiliary Variables in Sample Survey. *Journal of Statistical Computation and Simulation,* 89**,** 2316-2327.

Arnab, R. (2017). *Survey Sampling Theory and Applications,* Academic Press.

Basseville, M. (2013). Divergence Measures for Statistical Data Processing—an Annotated Bibliography. *Signal Processing,* 93**,** 621-633.

Basu, A., Shioya, H. & Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*, Chapman and Hall/CRC.

Bethlehem, J. G. & Keller, W. J. (1987). Linear Weighting of Sample Survey Data. *Journal of official Statistics,* 3**,** 141-153.

Bocci, J. & Beaumont, C. (2008). Another Look at Ridge Calibration. *Metron,* 66**,** 5-20.

Brewer, K. (1963). Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process. *Australian Journal of statistics,* 5**,** 93-105.

Brewer, K. R. & Hanif, M. (2013). *Sampling with Unequal Probabilities,* Springer Science & Business Media.

Cardot, H., Goga, C. & Shehzad, M.-A. (2017). Calibration and Partial Calibration on Principal Components When the Number of Auxiliary Variables Is Large. *Statistica Sinica***,** 243-260.

Cassel, C.-M., Sarndal, C.-E. & Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling,* Wiley.

Cassel, C. M., Särndal, C. E. & Wretman, J. H. (1976). Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. *Biometrika,* 63**,** 615-620.

Chambers, R. (2011). Which Sample Survey Strategy? A Review of Three Different Approaches. *Pakistan Journal of Statistics* 27(4)**,** 337-357.

Chambers, R. L. & Clark, R. G. (2008). Adaptive Calibration for Prediction of Finite Population Totals. *Survey Methodology,* 34(2)**,** 163-172.

Chen, J. & Qin, J. (1993). Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information. *Biometrika,* 80**,** 107-116.

Clement, E. P. & Enang, E. I. (2015). Calibration Approach Alternative Ratio Estimator for Population Mean in Stratified Sampling. *International Journal of Statistics and Economics,* 16**,** 83-93.

Cochran, W. (1942). Sampling Theory When the Sampling-Units Are of Unequal Sizes. *Journal of the American Statistical Association,* 37**,** 199-212.

Cochran, W. G. (1977). *Sampling Techniques: 3d Ed*, Wiley New York.

Cochran, W. G. (2007). *Sampling Techniques*, John Wiley & Sons.

Cressie, N. & Read, T. R. (1984). Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society: Series B (Methodological),* 46**,** 440-464.

Das, A. (1951). On Two Phase Sampling and Sampling with Varying Probabilities. *Bulletin of the International Statistical Institute,* 33**,** 105-112.

Deville, J.-C. & Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American statistical Association,* 87**,** 376-382.

Deville, J.-C. & Tillé, Y. (2004). Efficient Balanced Sampling: The Cube Method. *Biometrika,* 91**,** 893-912.

Duchesne, P. (1999). Robust Calibration Estimators. *Survey Methodology,* 25**,** 43-56.

Elabid, I. & Ismail, Z. (2017). A Comparison of Two Alternative Procedures for the Classical Scenario of Probability Proportional to Size. *Far East Journal of Mathematical Sciences (FJMS),* 102**,** 2531 - 2550.

Estevao, V. M. & Särndal, C.-E. (2002). The Ten Cases of Auxiliary Information for Calibration in Two-Phase Sampling. *Journal of Official Statistics*, 18**,** 233.

Farrell, P. J. & Singh, S. Penalized Chi Square Distance Function in Survey Sampling. ASA Proceedings, 2002. 963-968.

Franklin, S. & Walker, C. (2003). Survey Methods and Practices, Statistics Canada. *Social Survey Methods Division, Ottawa*.

Fuller, W. A. (2011). *Sampling Statistics*, John Wiley & Sons.

Godambe, V. (1955). A Unified Theory of Sampling from Finite Populations. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 269-278.

Goga, C. & Shehzad, M. A. (2014). A Note on Partially Penalized Calibration. *Pak. J. Statist*, 30**,** 429-438.

Grafström, A. (2010). *On Unequal Probability Sampling Designs.* Ph.D-thesis, Department of Mathematics and Mathematical Statistics, Umea university, Sweden.

Gregoire, T. G. (1998). Design-Based and Model-Based Inference in Survey Sampling: Appreciating the Difference. *Canadian Journal of Forest Research,* 28**,** 1429-1447.

Guggemos, F. & Tillé, Y. (2010). Penalized Calibration in Survey Sampling: Design-Based Estimation Assisted by Mixed Models. *Journal of statistical planning and inference,* 140**,** 3199-3212.

Hansen, M. H. & Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *The Annals of Mathematical Statistics,* 14**,** 333-362.

Hansen, M. H., Hurwitz, W. N. & Madow, W. G. (1953a). *Sample Survey Methods and Theory*, Wiley New York.

Hansen, M. H., Hurwitz, W. N. & Madow, W. G. (1953b). *Sample Survey Methods and Theory. Vol. 1, Methods and Applications,* Wiley.

Hanurav, T. (1966). Some Aspects of Unified Sampling Theory. *Sankhyā: The Indian Journal of Statistics, Series A***,** 175-204.

Hartley, H. & Rao, J. (1962). Sampling with Unequal Probabilities and without Replacement. *The Annals of Mathematical Statistics***,** 350-374.

Hartley, H. & Ross, A. (1954). Unbiased Ratio Estimators. *Nature*, 174**,** 270-271.

Holmberg, A. (2003). *Essays on Model Assisted Survey Planning.* Acta Universitatis Upsaliensis.

Horvitz, D. G. & Thompson, D. J. (1952). A Generalization of Sampling without Replacement from a Finite Universe. *Journal of the American statistical Association*, 47**,** 663-685.

Joshi, K. & Rajarshi, M. (2018). Modified Probability Proportional to Size Sampling. *Communications in Statistics-Theory and Methods,* 47**,** 805-815.

Juárez, S. F. & Schucany, W. R. (2006). A Note on the Asymptotic Distribution of the Minimum Density Power Divergence Estimator. *Lecture Notes-Monograph Series***,** 334-339.

Kim, J. K. (2009). Calibration Estimation Using Empirical Likelihood in Survey Sampling. *Statistica Sinica***,** 145-157.

Kim, J. K. & Park, M. (2010). Calibration Estimation in Survey Sampling. *International Statistical Review,* 78**,** 21-39.

Kott, P. S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology,* 32**,** 133.

Koyuncu, N. (2018). Calibration Estimator of Population Mean under Stratified Ranked Set Sampling Design. *Communications in Statistics-Theory and Methods,* 47**,** 5845-5853.

Koyuncu, N. & Kadilar, C. (2013). Calibration Estimator Using Different Distance Measures in Stratified Random Sampling. *International Journal of Modern Engineering Research,* 3**,** 415-419.

Koyuncu, N. & Kadilar, C. (2016). Calibration Weighting in Stratified Random Sampling. *Communications in Statistics-Simulation and Computation,* 45**,** 2267-2275.

Laaksonen, S. (2006). Need for High Quality Auxiliary Data Service for Improving the Quality of Editing and Imputation. *United Nations Statistical Commission,"Statistical Data Editing,* 3**,** 334-344.

Lindsay, B. G. (1994). Efficiency Versus Robustness: The Case for Minimum Hellinger Distance and Related Methods. *The annals of statistics,* 22**,** 1081-1114.

Madow, W. G. (1949). On the Theory of Systematic Sampling, Ii. *The Annals of Mathematical Statistics***,** 333-354.

Markatou, M., Chen, Y., Afendras, G. & Lindsay, B. G. (2017). Statistical Distances and Their Role in Robustness. *New Advances in Statistics and Data Science.* Springer.

Maronna, M. & Suggests, M. (2009). *Package 'Robustbase'* [Online]. Available: https://vincentarelbundock.github.io/Rdatasets/datasets.html [19 April 2009].

Mattheou, K. & Karagrigoriou, A. (2010). A New Family of Divergence Measures for Tests of Fit. *Australian & New Zealand Journal of Statistics,* 52**,** 187-200.

Menéndez, M., Morales, D., Pardo, L. & Salicrú, M. (1995). Asymptotic Behaviour and Statistical Applications of Divergence Measures in Multinomial Populations: A Unified Study. *Statistical Papers,* 36**,** 1-29.

Mickey, M. R. (1959). Some Finite Population Unbiased Ratio and Regression Estimators. *Journal of the American Statistical Association,* 54**,** 594-612.

Murthy, M. (1957). Ordered and Unordered Estimators in Sampling without Replacement. *The Indian Journal of Statistics (1933-1960),* 18**,** 379-390.

Murthy, M. (1964). Product Method of Estimation. *Sankhyā: The Indian Journal of Statistics, Series A***,** 69-74.

Narain, R. (1951). On Sampling without Replacement with Varying Probabilities. *Journal of the Indian Society of Agricultural Statistics,* 3**,** 169-174.

Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society,* 97**,** 558-625.

Pardo, L. (2018). *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC.

Parr, W. C. & De Wet, T. (1981). On Minimum Cramer-Von Mises-Norm Parameter Estimation. *Communications in Statistics-Theory and Methods,* 10**,** 1149-1166.

Prasad, B. (1989). Some Improved Ratio Type Estimators of Population Mean and Ratio in Finite Population Sample Surveys. *Communications in Statistics-Theory and Methods,* 18**,** 379-392.

Raj, D. (1956). Some Estimators in Sampling with Varying Probabilities without Replacement. *Journal of the American Statistical Association,* 51**,** 269-284.

Raj, D. (1956 a). Some Estimators in Sampling with Varying Probabilities without Replacement. *Journal of the American Statistical Association,* 51**,** 269-284.

Rao, J. (1963). On Three Procedures of Unequal Probability Sampling without Replacement. *Journal of the American Statistical Association,* 58**,** 202-215.

Rao, J. N., Hartley, H. & Cochran, W. (1962). On a Simple Procedure of Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 482-491.

Rao, T. (2015). The Neglected Das Estimator and Related Results. Unpublished manuscripts.

Robson, D. S. (1957). Applications of Multivariate Polykays to the Theory of Unbiased Ratio-Type Estimation. *Journal of the American Statistical Association*, 52**,** 511-522.

Roy, J. & Chakravarti, I. (1960). Estimating the Mean of a Finite Population. *The Annals of Mathematical Statistics***,** 392-398.

Salinas, V. I., Sedory, S. A. & Singh, S. (2019). Calibrated Estimators in Two-Stage Sampling. *Communications in Statistics-Theory and Methods,* 48**,** 1449-1469.

Samawi, H. M. & Muttlak, H. A. (1996). Estimation of Ratio Using Rank Set Sampling. *Biometrical Journal,* 38**,** 753-764.

Sampath, S. (2001). *Sampling Theory and Methods,* CRC press.

Särndal, C.-E. (1996). Efficient Estimators with Simple Variance in Unequal Probability Sampling. *Journal of the American Statistical Association,* 91**,** 1289-1300.

Särndal, C.-E. (2007). The Calibration Approach in Survey Theory and Practice. *Survey Methodology,* 33**,** 99-119.

Särndal, C.-E., Swensson, B. & Wretman, J. (2003). *Model Assisted Survey Sampling,* Springer Science & Business Media.

Särndal, C.-E., Thomsen, I., Hoem, J. M., Lindley, D., Barndorff-Nielsen, O. & Dalenius, T. (1978). Design-Based and Model-Based Inference in Survey Sampling [with Discussion and Reply]. *Scandinavian Journal of Statistics,* 5**,** 27-52.

Särndal, C., Swensson, B. & Wretman, J. (1992). Model Assisted Survey Sampling Springer. *New York*.

Sarndal, C. E. (1980). On Π-Inverse Weighting Versus Best Linear Unbiased Weighting in Probability Sampling. *Biometrika,* 67**,** 639-650.

Scott, A. & Wu, C.-F. (1981). On the Asymptotic Distribution of Ratio and Regression Estimators. *Journal of the American Statistical Association,* 76**,** 98-102.

Sen, A. R. (1953). On the Estimate of the Variance in Sampling with Varying Probabilities. *Journal of the Indian Society of Agricultural Statistics,* 5**,** 127.

Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael 'Selected' Amy,* Springer Science & Business Media.

Singh, S. & Arnab, R. (2006). A Bridge between the Greg and the Linear Regression Estimators. Joint Statistical Meeting, ASA Section on Survey Research Methods, Seattle. 3689-3693.

Singh, S. & Arnab, R. (2011). On Calibration of Design Weights. *Metron,* 69**,** 185-205.

Singh, S., Horn, S., Chowdhury, S. & Yu, F. (1999). Theory & Methods: Calibration of the Estimators of Variance. *Australian & New Zealand Journal of Statistics,* 41**,** 199-212.

Singh, S. & Sedory, S. A. (2016). Two-Step Calibration of Design Weights in Survey Sampling. *Communications in Statistics-Theory and Methods,* 45**,** 3510-3523.

Singh, S., Sedory, S. A., Rueda, M. D. M., Arcos, A. & Arnab, R. (2015). *A New Concept for Tuning Design Weights in Survey Sampling: Jackknifing in Theory and Practice,* Academic Press.

Srivenkataramana, T. & Tracy, D. (1980). An Alternative to Ratio Method in Sample Surveys. *Annals of the Institute of Statistical Mathematics,* 32**,** 111-120.

Stearns, M. & Singh, S. (2008). On the Estimation of the General Parameter. *Computational Statistics & Data Analysis,* 52**,** 4253-4271.

Stehman, S. V. & Overton, W. S. (1987). Estimating the Variance of the Horvitz-Thompson Estimator in Variable Probability Systematic Samples. Proceedings of the Section on Survey Research Methods. 743-748.

Subramani, J. (2013). Generalized Modified Ratio Estimator for Estimation of Finite Population Mean. *Journal of Modern Applied Statistical Methods,* 12**,** 7.

Thompson, M. (1997). *Theory of Sample Surveys,* CRC Press.

Tillé, Y. (2011a). *Sampling Algorithms,* New York., Springer.

Tillé, Y. (2011b). Ten Years of Balanced Sampling with the Cube Method: An Appraisal. *Survey methodology,* 37**,** 215-226.

Tiwari, N. & Chilwal, A. (2014). Empirical Comparison of Various Approximate Estimators of the Variance of Horvitz Thompson Estimator under Split Method of Sampling. *Journal of Reliability and Statistical Studies,* 7.

Traat, I., Bondesson, L. & Meister, K. (2000). *Distribution Theory for Sampling Designs,* Department of Mathematical Statistics, Ume° a University.

Tracy, D., Singh, S. & Arnab, R. (2003). Note on Calibration in Stratified and Double Sampling. *Survey Methodology,* 29**,** 99-104.

Upadhyaya, L. N. & Singh, H. P. (1999). Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences,* 41**,** 627-636.

Vajda, I. (1989). Theory of Statistical Inference and Information, Vol. 11. Kluwer Academic Pub.

William, C. P. (1981). Minimum Distance Estimation: A Bibliography. *Communications in Statistics-Theory and Methods,* 10**,** 1205-1224.

Wolfowitz, J. (1952). Consistent Estimators of the Parameters of a Linear Structural Relation. *Scandinavian Actuarial Journal,* 1952**,** 132-151.

Wolfowitz, J. (1953). Estimation by the Minimum Distance Method. *Annals of the institute of Statistical Mathematics,* 5**,** 9-23.

Wolfowitz, J. (1954). Estimation by the Minimum Distance Method in Nonparametric Stochastic Difference Equations. *The Annals of Mathematical Statistics,* 25**,** 203-217.

Wolfowitz, J. (1957). The Minimum Distance Method. *The Annals of Mathematical Statistics,* 28**,** 75-88.

Wright, T. & Farmer, J. (2000). *A Bibliography of Selected Statistical Methods and Development Related to Census 2000*, US Bureau of the Census.

Wu, C. & Lu, W. W. (2016). Calibration Weighting Methods for Complex Surveys. *International Statistical Review,* 84**,** 79-98.

Wu, C. & Sitter, R. R. (2001). A Model-Calibration Approach to Using Complete Auxiliary Information from Survey Data. *Journal of the American Statistical Association*, 96**,** 185-193.

Yates, F. & Grundy, P. (1953). Selection without Replacement from within Strata with Probability Proportional to Size. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 253-261.

## LIST OF PUBLICATIONS

## Journal Publications

Elabid, I. & Ismail, Z. (2017). A Comparison of Two Alternative Procedures for the Classical Scenario of Probability Proportional to Size. *Far East Journal of Mathematical Sciences (FJMS),* 102**,** 2531 - 2550.

## National Conference

Ibrahim Elabid , Z. I. (2015). Procedures for Estimating the Population Total in Unequal Probability Sampling Designs. *In Proceeding the Third International Science Postgraduate Conference (ISPC2015).* Faculty of Science, UTM Johor Bahru, Malaysia.

Ibrahim Elabid , Z. I. (2016). A Comparison of Alternative Procedures for the Classical Scenario for Estimating the Population Total. *The Third ISM International Statistical Conference* Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia.

Zuhaimy Ismail, I. E. (2014). Chronological Review on Unequal Probability Sampling Designs. *In Proceeding The Second International Science Postgraduate Conference (ISPC2014).* Faculty of Science, UTM Johor Bahru, Malaysia.