

**MODIFIED ITERATIVE AND REPETITIVE
ESTIMATION METHODS FOR FITTING
MEASUREMENT ERROR MODEL**

RO'YA SALEH FALEH AL-DIBI'I

UNIVERSITI SAINS MALAYSIA

2025

**MODIFIED ITERATIVE AND REPETITIVE
ESTIMATION METHODS FOR FITTING
MEASUREMENT ERROR MODEL**

by

RO'YA SALEH FALEH AL-DIBI'I

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

March 2025

ACKNOWLEDGEMENT

Alhamdulillah with the blessings and hidayah of Mighty Allah SWT throughout my life, finally I am able to complete my PhD Thesis. I would like to thank my Supervisor, Dr. Rosmanjawati Abdul Rahman and my Co-Supervisor Professor Dr. Amjad Al-Nasser, for pushing me throughout the most intellectually challenging journey I have ever taken. Without the guidance, support and motivation, the accomplishment of this thesis would not have been possible. I would not have had this opportunity without the consent of my supervisors. They have had a tremendous influence on my professional as well as personal development as my teachers, mentors and coaches. They helped me continuously during my PhD to grow as a researcher. I am also grateful to all my colleagues and friends at USM for their continuous encouragement and assistance throughout my PhD journey.

Lastly, I would be remiss in not mentioning my family, especially my mother. She belief in me has kept my spirits and motivation high during this process, also for her unconditional support and prayers throughout my educational career.

I am also grateful to my beloved mother, for her unconditional support and prayers throughout my educational career.

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
ABSTRAK	xi
ABSTRACT.....	xiii
CHAPTER 1 INTRODUCTION.....	1
1.1 Introduction to the Research Topic.....	1
1.2 Problem Statement.....	7
1.3 Research Objectives (RO) of the Study.....	9
1.4 Research Questions (RQ).....	9
1.5 Significance of the Study.....	10
1.6 Scope and Limitations.....	12
1.7 Organizations of the Thesis	14
CHAPTER 2 LITERATURE REVIEW	15
2.1 Overview.....	15
2.2 Problems related to measurement error model (MEM).....	15
2.3 Existing approaches to fitting MEM.....	19
2.3.1 Instrumental variable estimation	20
2.3.2 Robust Theil-Type Estimators.....	22
2.3.3 Maximum Likelihood Estimation method.....	23
2.3.4 Other Estimation Approaches.....	25

2.4	Challenges in MEM	29
2.5	Concluding Remarks.....	32
CHAPTER 3 RESEARCH METHODOLOGY		33
3.1	Overview	33
3.2	The Estimation Methods of a Simple Measurement Error Model.....	34
3.2.1	Classical Methods.....	35
3.2.1(a)	Maximum Likelihood Estimation Method.....	35
3.2.1(b)	Method of Moments	38
3.2.1(c)	Wald-Type Estimation Method	39
3.2.2	The New Proposed Estimation Methods	41
3.2.2(a)	Repetitive Estimation Procedure	42
3.2.2(b)	Iterative Procedures	42
3.2.3	Simulation Algorithm	44
3.3	The Estimation Methods of Multiple Measurement Error Model	46
3.3.1	Classical Estimation Methods for Fitting Multiple MEM.....	46
3.3.1(a)	Maximum Likelihood Estimator	46
3.3.1(b)	Method of Moments	48
3.3.2	The New Weighted Grouping Method	49
3.3.2(a)	Weighted Wald-Type Procedures.	51
3.3.2(b)	Repetitive Weighted Grouping Procedure	53
3.3.2(c)	An Iterative Weighted Procedure	54
3.3.3	Weighted Robust Method.....	56
3.3.3(a)	Modified Theil Estimator	58
3.3.3(b)	Modified Siegel Estimator.....	59
3.4	Simulation Algorithm for Multiple MEM	60

3.5	Summary of Methodology	64
CHAPTER 4 SIMULATIONS RESULTS		65
4.1	Overview	65
4.2	Fitting a Simple MEM	65
4.3	Fitting a Multiple MEM.....	71
4.3.1	Classical Methods.....	71
4.3.2	The Weighted Grouping Method.....	74
4.3.3	Weighted Robust Method.....	87
4.4	Summary	94
CHAPTER 5 EMPIRICAL RESULTS FROM DATASETS		96
5.1	Overview	96
5.2	Dataset 1: Data of Infant Deaths, Hepatitis B and Polio.....	96
5.2.1	Data Overview	96
5.2.2	Summary Statistics	98
5.2.3	Modeling Infant Deaths and Hepatitis B Data.....	100
5.2.4	Modeling Infant Deaths, Hepatitis B and Polio.....	102
5.2.4(a)	Weighted Grouping Methods	102
5.2.4(b)	Weighted Robust Methods	105
5.3	Dataset 2: Data of HDI, GDP and Unemployment Rate of Jordan	107
5.3.1	Data Overview	108
5.3.2	Summary Statistics	110
5.3.3	Modeling HDI and GDP Data	112
5.3.4	Modeling HDI, Unemployment Rate and GDP Data	114
5.3.4(a)	Weighted Grouping Methods	114
5.3.4(b)	Weighted Robust Methods	116

5.4	Summary	119
CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS		121
6.1	Overview	121
6.2	Summary of Key Findings	122
6.3	Contributions to the Study	123
6.4	Limitations and Suggestions for Future Research	124
REFERENCES.....		125
LIST OF PUBLICATIONS		

LIST OF TABLES

		Page
Table 2.1	MLE of the slope for normal structural MEM (Cheng & Ness, 1999)	25
Table 3.1	Average of Bias and MSE for α , β_1 and β_2 for sample size (200) without outlier, (random samples=10000)..	62
Table 4.1	Average of Bias and MSE of α and β for samples without outlier.	67
Table 4.2	Average of Bias and MSE for α and β when $\sigma\delta^2 = 16$ with outliers in x.	68
Table 4.3	Average of Bias and MSE for α and β when $\sigma\epsilon^2 = 16$ with outliers in y.	69
Table 4.4	Average of Bias and MSE for α and β when $(\sigma\epsilon^2, \sigma\delta^2) = (16, 16)$ with outliers in both (x,y).	70
Table 4.5	Average of Bias and MSE of α , β_1 and β_2 of MLE and MOM with under varying outlier conditions	72
Table 4.6	Average of Bias and MSE of α and β of weighted grouping method for samples without outlier.	75
Table 4.7	Average of Bias and MSE for α and β of weighted grouping method for samples when $\sigma\delta_{12}^2 = 16$ with outliers in x_1	77
Table 4.8	Average of Bias and MSE for α and β of weighted grouping method for samples when $\sigma\delta_{22}^2 = 16$ with outliers in x_2	79
Table 4.9	Average of Bias and MSE for α and β of weighted grouping method for samples when $(\sigma\delta_{12}^2, \sigma\delta_{22}^2) = (16, 16)$ with outliers in both (x_1, x_2)	81
Table 4.10	Average of Bias and MSE for α and β of weighted grouping method for samples when $\sigma\epsilon^2 = 16$ with outliers in y	83
Table 4.11	Average of Bias and MSE for α and β of weighted grouping method for samples when $(\sigma\delta_{12}^2, \sigma\delta_{22}^2, \sigma\epsilon^2) = (16, 16, 16)$ with outliers in all (x_1, x_2, y)	85
Table 4.12	Average of Bias and MSE of α and β of weighted robust method for samples or samples without outlier.	88

Table 4.13	Average of Bias and MSE for α and β of weighted robust method for samples when $\sigma\delta_{12} = 16$ with outliers in x_1	89
Table 4.14	Average of Bias and MSE for α and β of weighted robust method for samples when $\sigma\delta_{22} = 16$ with outliers in x_2	90
Table 4.15	Average of Bias and MSE for α and β of weighted robust method for samples when $(\sigma\delta_{12}, \sigma\delta_{22}) = (16, 16)$ with outliers in both (x_1, x_2)	91
Table 4.16	Average of Bias and MSE for α and β of weighted robust method for samples when $\sigma\epsilon_2 = 16$ with outliers in y	92
Table 4.17	Average of Bias and MSE for α and β of weighted robust method for samples when $(\sigma\delta_{12}, \sigma\delta_{22}, \sigma\epsilon_2) = (16, 16, 16)$ with outliers in all (x_1, x_2, y)	93
Table 5.1	Descriptive Statistics.....	98
Table 5.2	Correlation Matrix between the dependent variable (infant deaths) and independent variables (Hepatitis B and Polio).	98
Table 5.3	Parameter Estimation of Infant Deaths Vs Hepatitis B.....	101
Table 5.4	Parameters Estimation of Infant Deaths, Hepatitis B and Polio	104
Table 5.5	Parameters Estimation of Infant Deaths, Hepatitis B and Polio	106
Table 5.6	Yearly Dataset of HDI, GDP and Unemployment Rate of Jordan (1990–2021)	109
Table 5.7	Descriptive Statistics.....	110
Table 5.8	Correlation Matrix between the dependent variable (HDI) and independent variables (GDP and Unemployment rate).	111
Table 5.9	Parameters Estimation of HDI vs GDP	113
Table 5.10	Parameters Estimation of HDI vs GDP and Unemployment Rate	115
Table 5.11	Parameters Estimation of HDI vs GDP and Unemployment Rate	118

LIST OF FIGURES

		Page
Figure 3.1	All possible slopes between the subgroups.....	42
Figure 3.2	Pairwise slope between the subgroups.....	43
Figure 3.3	Formulas for computing the new weighted estimators	53
Figure 3.4	All possible slopes between the subgroups.....	54
Figure 3.5	An Illustration of the Pairwise Slope Between the Subgroups	55
Figure 5.1	The line plot of Infant Deaths and Hepatitis B	99
Figure 5.2	The line plot of Infant Deaths and Polio	99
Figure 5.3	The scatter plot of infant deaths, Hepatitis B and Polio	99
Figure 5.4	Residual plots of the estimation methods	102
Figure 5.5	Residual plots of the estimation methods for Weight Case 1	104
Figure 5.6	Residual plots of the estimation methods for Weight Case 2	105
Figure 5.7	Residual plots of the estimation methods for Weight Case 1	107
Figure 5.8	Residual plots of the estimation methods for Weight Case 2	107
Figure 5.9	The line plot of HDI and Unemployment rate	111
Figure5.10	The line plot of HDI and GDP	111
Figure 5.11	The scatter plot of HDI, GDP and unemployment rate.....	111
Figure 5.12	Residual plots of the estimation methods	113
Figure 5.13	Residual plots of the estimation methods for Weight Case 1	116
Figure 5.14	Residual plots of the estimation methods for Weight Case 2	116
Figure 5.15	Residual plots of the estimation methods for Weight Case 1	118
Figure 5.16	Residual plots of the estimation methods for Weight Case 2	118

LIST OF ABBREVIATIONS

IWM	Iterative Weighted Method
MEM	Measurement Error Model
MLE	Maximum Likelihood Estimation
MOM	Method of Moment
MSE	Mean Square Error
SMEM	Structural Measurement Error Models
WGM	Weighted Grouping Method

KAEDAH PENGANGGARAN LELARAN DAN ULANGAN TERUBAH UNTUK PENYUAIAN MODEL RALAT SUKATAN

ABSTRAK

Tesis ini memperkenalkan kaedah penganggaran baharu untuk penyuaian Model Ralat Sukatan (MEM) berstruktur, yang menangani cabaran dalam menganggar hubungan apabila semua pemboleh ubah tertakluk kepada ralat. Kajian ini memberi tumpuan kepada teknik penganggaran berpemberat yang bebas taburan, yang meningkatkan penganggaran parameter tanpa memerlukan andaian ketat seperti yang diperlukan oleh kaedah tradisional. Prosedur penganggaran Wald berpemberat baharu, yang dibangunkan dalam bentuk lelaran dan ulangan terubah, dicadangkan dan diuji pada MEM mudah dan berganda. Kajian simulasi menunjukkan bahawa kaedah ini mencapai ralat min kuasa dua yang lebih rendah merentasi pelbagai senario pemberat, mengatasi kaedah klasik seperti Penganggaran Kebarangkalian Maksimum (MLE) dan Kaedah Momen (MOM) dari segi ketepatan dan kebolehpercayaan. Selain itu, kaedah penganggaran berpemberat yang teguh, termasuk prosedur Theil dan Siegel, menunjukkan prestasi yang lebih baik dalam pelbagai keadaan. Teknik yang dicadangkan ini diterapkan pada dua set data sebenar: pertama, mengkaji hubungan antara kematian bayi, Hepatitis B, dan Polio yang menunjukkan korelasi negatif yang kuat, dan kedua, menganalisis hubungan antara Keluaran Dalam Negara Kasar (KDNK), pengangguran, dan Indeks Pembangunan Manusia (HDI) di Jordan, yang menunjukkan korelasi positif yang kuat antara KDNK dan HDI serta hubungan negatif yang kuat antara pengangguran dan HDI. Kaedah ini, disokong oleh alat pengiraan baharu yang dibangunkan dalam tesis ini, memberikan penambahbaikan yang

signifikan dalam penyuaian MEM dan menawarkan rujukan yang bernilai untuk penyelidikan masa depan dalam analisis data multivariat.

**MODIFIED ITERATIVE AND REPETITIVE ESTIMATION METHODS
FOR FITTING MEASUREMENT ERROR MODEL**

ABSTRACT

This thesis introduces new estimation methods for fitting structural Measurement Error Model (MEM), addressing challenges in estimating relationships where all variables are subject to error. The study focuses on distribution-free weighted estimation techniques, which improve parameter estimation without the strict assumptions required by traditional methods. A new weighted Wald-type estimation procedure, developed in modified iterative and repetitive forms, is proposed and tested on both simple and multiple MEMs. Simulation studies show that these methods achieve lower mean square errors across various weighting scenarios, outperforming classical approaches such as Maximum Likelihood Estimation (MLE) and the Method of Moments (MOM) in terms of accuracy and reliability. Additionally, robust weighted estimation methods, including Theil and Siegel procedures, demonstrate superior performance in different settings. The proposed techniques are applied to two real datasets: one examining the relationship between infant deaths, Hepatitis B, and Polio, revealing strong negative correlations, and another analyzing Gross Domestic Product (GDP), unemployment, and the Human Development Index (HDI) in Jordan, highlighting strong positive correlations between GDP and HDI and a strong negative relationship between unemployment and HDI. These methods, supported by new computational tools developed in this thesis, provide significant improvements in MEM fitting and offer a valuable reference for future research in multivariate data analysis.

CHAPTER 1

INTRODUCTION

1.1 Introduction to the Research Topic

Accurately determining the interconnections among variables in statistical analysis is crucial for deciphering complex phenomena across numerous scientific fields. Fundamental to this exploration are linear regression models, both in their simple and multiple variants, which are instrumental in analyzing these links, forecasting future events, and guiding decision-making. Nonetheless, the effectiveness of these models is contingent upon the precise measurement of variables. The issue of measurement error, wherein the variables of interest are inaccurately captured, undermines the validity of statistical conclusions, leading to skewed and unreliable estimates of parameters (Fuller, 1987; Cheng & Van Ness, 1999). It is vital, therefore, to acknowledge and rectify measurement errors to preserve the trustworthiness of research outcomes, which supports both theoretical progress and practical implementations.

This research delves into the Measurement Error Model (MEM), a strategic approach compensating for inaccuracies in measuring variables within regression analysis. The MEM posits that additive mistakes often conceal the authentic values of variables; thus, they are only indirectly discernible through their observed equivalents. By adopting standard linear MEMs and their advancements, such as Structural Measurement Error Models (SMEM), the study seeks to correct the distortions introduced by measurement errors using intricate statistical methods (Pagan & Pesaran, 2008; Fuller, 1987). However, the success of these rectification techniques heavily depends on the assumptions they make, which, if breached, can significantly

impair their efficacy. This challenge is exacerbated when it is difficult or unfeasible to validate assumptions concerning the distribution of error terms with the data (Spicker et al., 2021).

Highlighting the paramount role of accurately correcting measurement errors for the soundness of statistical analyses and the confines of existing methods, this study introduces a Modified Iterative Estimation Method that involves a weighted grouping method and a weighted robust method to improve the fit of MEMs. By overcoming the limitations faced by classical estimation techniques, which falter amid non-normal distributions of hidden variables and when the foundational presumptions of conventional correction methods are unmet, this investigation strives to provide a more dependable and effective solution for addressing measurement errors in linear regression models. The imperative for such developments emphasizes the overarching relevance of this research field, as the pursuit of precision in measuring variables extends beyond specific disciplines, influencing the credibility of scientific exploration and the success of policies and practices rooted in evidence.

In analyzing relationships between two or more variables, researchers use a simple linear regression when there is one dependent variable and one explanatory variable; otherwise, a Multiple Linear Regression is used when there are several explanatory variables. The basic assumption on the explanatory variable is that the researcher measures them without any error. If this assumption is violated, then the model is known as the Measurement Error Model (MEM) (Fuller, 1987; Kendall & Stuart, 1979). It is worth saying that ignoring measurement errors in analyzing relationships when they exist can cause misleading results, such as bias and

inconsistency in parameter estimation (Cheng & Van Ness, 1999). The corresponding standard linear MEM assumes that two variables ξ and η are related by:

$$\eta = \alpha + \beta_1 \xi \quad (1.1)$$

However, the variables ξ and η are unobservable and can only be observed with additive errors. Thus, instead of observing ξ and η directly, one observes the variables

$$x = \xi + \delta \quad (1.2)$$

$$y = \eta + \varepsilon \quad (1.3)$$

where ξ and the errors, δ and ε , are uncorrelated. A sample of size n for this model would be

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix} \quad (1.4)$$

Since both ξ and η are measured with error, this model is called the MEM. We call this the ‘standard’ MEM. The variables ξ and η are sometimes also called latent variables (Carroll et al., 1995). Considering a sample size n , the linear univariate MEM can be formulated as follows:

$$\eta = \alpha + \beta \xi \quad , \quad i = 1, 2, \dots, n \quad (1.5)$$

However, one observes (x_i, y_i) , $i = 1, 2, \dots, n$, which are the true variables plus the additive errors, $(\delta_i, \varepsilon_i)$:

$$x_i = \xi_i + \delta_i, \quad i = 1, 2, \dots, n \quad (1.6)$$

$$y_i = \eta_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1.7)$$

The linear model in Equation (1.5) assumes that the response variable y is not accurately measured while the predictor variable is perfectly measured. Alternatively, to be more convenient in modelling the relationship between two variables, one can use the Structural Measurement Error Model (SMEM) as an extension of the simple linear regression model by assuming both variables (response and predictor) are measured with independent errors (Adusumilli & Otsu, 2018; Fuller, 1987).

Measurement error, occurring when a variable of interest is not accurately observed, is a widespread problem that frequently calls into question the analysis's reliability. Although there are numerous methods for correcting the effects of measurement error, they become unreliable when their underlying assumptions are violated. This issue is exacerbated when presumptions, like those about the distribution of error terms, are hard or impossible to test using the available data. In literature, modelling the MEM frequently assumes an additive model with normally distributed errors. Despite being simple and appealing, this assumption must be corrected in practical applications, as discussed in Spicker et al. (2021). However, it is an extension of the simple linear regression model by assuming both variables (response and predictors) are measured with independent errors. The multiple measurement error models with vector-valued explanatory variables with more than one x variable. Consequently, the corresponding standard linear MEM Cheng and Van Ness (1999) is presented in Equations (1.2) and (1.3) as:

$$\eta_i = \alpha + \beta_{i1}\xi_{i1} + \beta_{i2}\xi_{i2} + \dots + \beta_{ik}\xi_{ik} \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k \quad (1.8)$$

$$y_i = \eta_i + \epsilon_i \quad (1.9)$$

$$x_{ij} = \xi_{ij} + \delta_{ij} \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k \quad (1.10)$$

where:

$$\epsilon_i = y_i - \eta_i \quad (1.11)$$

$$\xi_{ij} = x_{ij} - \delta_{ij} \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k \quad (1.12)$$

η_i , ϵ_i and δ_{ij} are mutually independent unobservable disturbances terms with zero means and finite variances. They are assumed to be independent of (ξ_{ij}, η_i) .

The MEM has a long history, and in the literature, two types of MEM are mentioned: structural MEM and functional MEM. The Structural Measurement Error Model (SMEM) analyses the structure of measurement errors and their association with the variable of interest. It employs a specific set of equations to depict how the actual values of variables relate to their observed measurements, factoring in the presence of errors during the measurement process (Bollen, 1989; Karlson et al., 2012). On the other hand, the Functional MEM is primarily concerned with defining the functional relationship between the actual values of the variables and their observed measurements. Unlike structural models, functional models do not explicitly address the mechanisms behind measurement errors. Instead, they focus on establishing a mathematical representation of the connection between the actual and measured values (Kiefer & Vogelsang, 2005; Carroll et al., 2006).

The independent variables in the functional relationships model are assumed to be fixed. In contrast, the independent variables in the structural relationships model are assumed to be random, with the same distribution across different populations. We assume that the distributions of independent variables in different populations have different location parameters in the ultrastructural case. The multivariate linear ultrastructural relationships model is the name given to this model. This model

includes the functional relationships model as a particular case because we only consider changes in location parameters.

Three separate MEM depend on the assumptions about ξ (Cheng & Van Ness, 1999):

- (1) If the ξ_i 's are unknown constants, then the model is known as a functional model.
- (2) If the ξ_i 's are independent identically distributed random variables and independent of the errors, then the model is known as a structural model, when $E(\xi_i) = \mu$ and $var(\xi_i) = \sigma^2$.
- (3) If the ξ_i 's are independent random variables but not identically distributed, then the model is known as ultrastructural model, when $E(\xi_i) = \mu$ and $var(\xi_i) = \sigma^2$.

Classical estimation methods, such as maximum likelihood, estimate the parameters of MEM when additional prior information is considered. However, challenges associated with using MLE, as noted by Chen and Van Ness (1999), Fuller (1987), and Carroll et al. (2006), have led researchers to explore alternative estimation procedures. To address these issues, this research proposes an iterative estimation procedure for fitting the MEM.

The main problem in Equation (1.5) is to estimate the unknown parameters α and β . Comprehensive reviews can be found in Cao et al. (2018). Generally, the variables δ and ε are assumed to be independent and normally distributed. However, there are situations when the proper distribution of the latent variable ξ departs from normality; that is, when skewness, outliers and multimodality are present. Then,

choosing more flexible models can be a valuable alternative to the normal one to overcome drawbacks (Cabral et al., 2020).

1.2 Problem Statement

This research targets the inherent difficulties in precisely fitting measurement error models (MEMs), primarily using classical estimation methods. MEMs play a pivotal role in statistical analyses by accurately depicting the relationship between variables, which may be affected by measurement errors. Classical estimation methods, like maximum likelihood estimation (MLE), are built on strict assumptions regarding error distributions and the non-existence of measurement errors conditions that often do not hold in practical methods. Dependence on these assumptions can result in biased and inconsistent parameter estimates, thus compromising the validity and reliability of the statistical conclusions drawn from such models.

The challenge is further amplified in the context of multiple MEMs, where vector-valued explanatory variables introduce additional complexity to the estimation process. Most prior studies have concentrated on MEMs with a single explanatory variable, primarily due to the complex challenges and the requirement for normal distribution assumptions in methods involving multiple MEMs. Even with this, the presumption of normally distributed errors, despite facilitating the estimation process, is not consistently applicable in empirical studies, leading to possible parameter misestimation and erroneous conclusions. Many authors have been discussed several estimation methods to fit the relationship between two variables (one dependent and one independent) when both are subject to error.

Numerous authors have discussed various estimation methods for modeling the relationship between two variables (one dependent and one independent) when both are subject to error. Among them are Al-Nasser (2010), Cao et al. (2018), Al-Nasser (2011, 2012), Schennach and Hu (2013), and Ghapor et al. (2015). However, relatively few studies have focused on multiple MEMs due to their complexity and reliance on normality assumptions, along with several other preconditions required to solve the model (Al-Nasser, 2014; Al-Nasser et al., 2016; Wiedermann et al., 2018; Adusumilli & Otsu, 2018; Schennach, 2013; Al-Nasser et al., 2005).

Furthermore, there needs to be more investigation into nonparametric and robust estimation methods that offer a more flexible and less assumption-dependent alternative to classical parametric methods. This cavity is particularly evident in multiple MEM methods, where pre-existing distributional and other assumptions significantly restrict the effectiveness and applicability of current estimation strategies.

This research proposes innovative new estimation methods to enhance the precision and reliability of parameter estimates in both simple and multiple MEMs. These methods are designed to minimize reliance on restrictive error distribution assumptions and to accommodate the complexities present in real data. By introducing this methods, the research fills a significant void in existing literature. It contributes to advancing more robust and efficient statistical analysis tools in the presence of measurement errors. This study aims to refine the methodological approach to MEM estimation, providing substantial implications for empirical research across various fields where accurate variable measurement and analysis are critical.

1.3 Research Objectives (RO) of the Study

The main objective of this study is to propose new procedures as tools for estimating parameters in simple and multiple MEMs. Specifically, this study aims to:

RO1. Derive the mathematical properties of the proposed estimators.

RO2. Conduct performance analysis and comparisons between the proposed methods and existing classical methods.

RO3. Apply the proposed procedures to model real-world datasets.

1.4 Research Questions (RQ)

In line with the problem identified and the primary goal of this investigation, the following research questions have been carefully crafted to steer the study towards an in-depth evaluation of this thesis. These inquiries aim to probe both the theoretical framework and the practical utility of the suggested estimation technique, examining its efficiency, dependability, and relevance in contrast to classical methods. The research questions are delineated as follows:

RQ1. What are the mathematical properties of the proposed estimators for simple and multiple MEMs, and how do these properties compare to those of classical estimation methods?

RQ2. How do the proposed estimation methods perform in terms of accuracy and reliability when compared to existing classical methods, particularly under different simulation scenarios?

RQ3. Can the proposed procedures effectively model real-world datasets, and how do the results compare to those obtained using classical estimation techniques?

These research questions are pivotal for deepening our comprehension of MEMs and augmenting the methodological arsenal available to researchers grappling with the pervasive issue of measurement error. Through meticulous exploration of these queries, this investigation aims to contribute to the statistical analysis domain, presenting innovative methods that enhance the accuracy and integrity of research outcomes amid measurement inaccuracies.

1.5 Significance of the Study

The significance of this investigation stems from its holistic strategy aimed at overcoming the widespread challenges associated with estimating Measurement Error Models (MEMs). Through developing and accessing a Modified New Estimation Method, this study seeks to bridge a significant void in the current discourse on statistical analysis in the context of measurement errors. The critical role of accurately modelling MEMs is emphasized by their extensive application across various scientific fields, where the measurement precision of variables is directly linked to the validity of research outcomes.

An essential contribution of this research is introducing an innovative estimation method that demonstrates resilience against the limitations commonly encountered with classical estimation methods. Classical methods often rely heavily on assumptions regarding the distribution of error terms and the non-existence of measurement errors, which can drastically constrain their utility and effectiveness. By proposing an alternative that reduces reliance on such assumptions, the study presents

a notable methodological improvement, increasing the accuracy and dependability of parameter estimates within MEMs.

Additionally, this research is distinguished by its attention to simple and multiple MEM contexts, an area that could be more extensively addressed in existing studies. The complexity and the involvement of vector-valued explanatory variables in multiple MEMs present significant estimation challenges. The new estimation methods introduced herein is crafted to manage these complexities adeptly, providing a versatile and less assumption-dependent solution suitable for various research contexts.

The practical ramifications of this study are extensive. In disciplines where measurement errors are common and can significantly influence the validity of findings, accessing a more trustworthy estimation methodology enhances the toolkit of researchers and analysts. This progress augments the precision of statistical analyses and enlarges the range of researchable questions, contributing to the expansion and depth of scientific exploration.

Moreover, this study enriches the theoretical underpinnings of statistical modelling amidst measurement inaccuracies by tackling the deficiencies of current estimation approaches and suggesting a methods adaptable to both simple and multiple MEMs. It meticulously evaluates the mathematical characteristics and efficacy of the Modified New Estimation Methods, offering insights into its applicability and constraints.

In essence, the value of this study lies in its dual contribution to the theoretical and practical aspects of statistical analysis involving MEMs. By introducing a pioneering estimation method that promises enhanced precision, reliability, and

adaptability, the research addresses a pivotal necessity within the field. It fills a notable gap in scholarly literature, advancing understanding of MEM estimation, and provides practical tools that could be utilized across various disciplines to improve the validity and credibility of research outcomes in scenarios plagued by measurement errors.

1.6 Scope and Limitations

The primary focus of this study is the development of a new procedure tools and their performance assessment tailored for fitting Measurement Error Models (MEMs) within contexts of both single and multiple variables. This research aims to delve into the mathematical attributes, foundational theories, and empirical performance of the proposed method compared to existing methods. It incorporates a thorough exploration via simulation studies and the application to real datasets to evaluate its efficiency, dependability, and suitability across various fields where measurement inaccuracies are prevalent.

Additionally, this study seeks to enhance the current knowledge base by tackling the drawbacks of classical estimation methods, especially their reliance on certain assumptions regarding error distributions and measurement accuracy. By offering an alternative solution that lessens these dependencies, the research intends to provide a versatile and resilient tool for researchers dealing with measurement error challenges in their datasets.

Nevertheless, several intrinsic limitations to the research design and methodology might influence the interpretation and extrapolation of the results. Initially, the validity and reliability of the new estimation methods are examined within the confines of the simulated data and specific real datasets selected for this study,

meaning the outcomes might differ under alternate datasets or conditions not examined herein.

Moreover, while the study strives for a detailed evaluation of the proposed method, the breadth of simulation studies and real data applications is restricted by practical constraints such as computational capabilities and data accessibility. These limitations may curtail the comprehensive exploration of the method's performance under every conceivable scenario and condition.

Thirdly, given the study's emphasis on MEMs, its findings are pertinent to statistical models and research instances where measurement errors are deliberately recognized and modelled. Consequently, the research's relevance and implications might be diminished when measurement inaccuracies are not the focal point or are managed through different methodological strategies.

Lastly, akin to all estimation methodologies, the proposed Modified Iterative Estimation Method operates on certain presumptions and mathematical constructs that delineate its functional framework. Although designed to counteract some common constraints of classical estimation methods, it may harbor limitations and assumptions, which could affect its appropriateness and efficacy in particular methods or with specific data types.

In essence, this research is characterized by its endeavor to refine the methodology for estimating MEMs through innovative estimation procedure. Despite its potential to significantly impact the field, the study's scope and influence are bounded by limitations concerning data selection, computational resources, specific application contexts, and the inherent traits of the proposed methodology. Recognizing

these limitations is vital for accurately interpreting the results and comprehending their implications for subsequent research and practical implementations.

1.7 Organizations of the Thesis

This thesis is divided into six detailed chapters, focusing on exploring measurements error models (MEM). The first chapter sets the stage for the research, establishing a foundational framework by detailing the motivation behind the study, presenting the problem statement, outlining the research objectives and scope, and emphasizing the study's contributions. Chapter 2 ventures into the historical backdrop, offering a chronological exploration of the literature on MEM, laying a robust groundwork for comprehending the current research landscape in this domain. The third chapter transitions to examining pertinent theories and methodologies related to MEM, furnishing the theoretical basis for the study. Chapter 4 is allocated to showcase the simulation outcomes of innovative estimation techniques for simple and multiple case analyses, illustrating the practical implementation of previously discussed theoretical concepts. Chapter 5 further enriches the investigation by presenting numerical examples that apply these novel estimation methods to real datasets, reinforcing the research findings' validity. Finally, Chapter 6, brings the thesis to a conclusion by critically evaluating the results and insights derived from the study, proposing recommendations for future research, and outlining the study's implications within the broader spectrum of MEMS research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter discusses the relevant literature for this thesis. There are six sections, including Section 2.2 which introduces the Measurement Error Model (MEM). Section 2.3 describes the existing approaches to fitting the MEM. Additionally, Section 2.4 highlights challenges in fitting the MEM. Section 2.5 covers some extensions from the existing methods in fitting the MEM, and finally, Section 2.6 summarizes this chapter.

2.2 Problems related to measurement error model (MEM)

MEM are extension models to the classical or ordinary linear regression models, either simple or multiple models. However, the MEM is used when the explanatory variable(s) is measured with error, such considerations lead to have incorrect parameter estimation. One of the most differences between both models (Linear regression and MEM), concerns the model identifiability. This problem occurs in the MEM, which means different set of the parameters could lead to the same joint distribution of the response and explanatory variables (Cheng & Van Ness, 1999). To clarify more differences between both models lets consider the mathematical relationship between two variables when both variables are subject to error, which is also known by the MEM with no equation error:

$$\eta = \alpha + \beta\xi \tag{2.1}$$

such that

$$y = \eta + \varepsilon$$

$$x = \xi + \delta$$

Then the differences between the linear regression model and the MEM as given by Hussin (1997) can be viewed as follows:

- (i) For any pair of observations (ξ_i, η_i) , ordinary linear regression assumes only the dependent variable η is observed with errors, and the independent variable ξ is free of error, whereas in the Measurement error models both variables are observed with errors.
- (ii) The MEM is symmetric in the roles played by η and ξ ; there is no distinction between "explanatory" and "response" variables or dependent and independent variables, unlike in ordinary linear regression.
- (iii) In ordinary replicated $\{(\xi_i, \eta_{ij}) : i = 1, 2, \dots, m, j = 1, 2, \dots, ni\}$ linear regression, any observation (ξ_i, η_{ij}) , has each η_{ij} paired with the common ξ_i , but in replicated MEM with data $\{(\xi_{ij}, \eta_{ik}) : i = 1, 2, \dots, m, j = 1, 2, \dots, ni, k = 1, 2, \dots, rj\}$, observations (ξ_{ij}) , and (η_{ik}) are not paired and there may even be different of ξ 's and η 's.
- (iv) Ordinary linear regression is more appropriate if our aim is to predict one variable from the other rather than to look at the underlying relationship between the two variables. Even if the ξ 's are measured with errors it may be appropriate to condition on the observed values of ξ 's when predicting η 's and use ordinary regression methods.

In fact, considering the simple linear relationship when the explanatory variable is measured with some errors, then the slope parameter is biased toward zero which is known by attenuation bias (Fuller, 1987). To clarify this problem, rewrite Equation (2.1) in the form of one model equation as:

$$y = \alpha + \beta (x - \delta) + \varepsilon$$

which can be simplify to:

$$y = \alpha + \beta x + u; \text{ where } u = \varepsilon - \delta x \quad (2.2)$$

Now, under the regular conditions of the classical regression model, which assumes x and u are independent, then for a given random sample of size n , $\{(x_i, y_i): i = 1, 2, \dots, n\}$; the ordinary least square (OLS) estimator of the slope parameter β will be:

$$\hat{\beta} = \frac{\sum_{j=1}^n (x_j - \bar{x}) y_j}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

However, using the MEM the explanatory variable x and the error term u are not independent and they are correlated with $Cov(x, u) = -\beta \sigma_\xi^2$. Therefore, the OLS estimator is inconsistent with probability limits equal to:

$$P(\lim \hat{\beta}) = \beta \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\xi^2}$$

Therefore, the OLS of the slope parameter is inconsistent with attenuation bias. Also, estimation of MEM becomes complicated as the numbers of unknowns exceeds the data points (this problem is very clear in the functional and ultra-structural MEMs), and this leads to an ill-posed problem. There are $(n+4)$ parameters to be estimated,

which are $\alpha, \beta, \sigma_{\xi}^2, \sigma_{\eta}^2$ and the latent variables or the so-called “incidental” parameters $\xi_1, \xi_2, \dots, \xi_n$ (or equivalently $\eta_1, \eta_2, \dots, \eta_n$). The presence of the incidental parameter leads to inconsistencies of the estimators (Kendall & Stuart, 1979).

This makes the traditional methods such as the maximum likelihood estimation (MLE) and OLS meaningless in estimating the model without the help of extra information. The problem with using OLS for fitting the MEM is that one must deal with two types of error within the minimization process in different directions the horizontal direction as well as the vertical direction. Unfortunately, the OLS mechanism required us to minimize these two types of errors symmetrically, which required to perform weighting of these two types of errors. This weighting is usually determined based on the ratio of the variances of the two errors. Moreover, solving the MEM problem using the MLE method requires the normality assumption for the two error terms. Solari (1969) showed that the maximum likelihood function does not give a maximum solution for the problem, instead it gives a saddle point. Therefore, additional assumption is required such as the information about the precision ratio to solve the problem.

The difficulty in using estimation methods (MLE and OLS) is that the ratio of the variances (the precision ratio) is not usually known and cannot usually be estimated. Alternatively, in the literature, there are two types of estimation procedures described to deal with the MEM, namely, weighted regression and instrumental variables.

In using the weighted regression required to pre-estimate the ratio of the variances of the two errors terms. While the Instrumental variable method concerns with obtaining a new “instrumental” variable that is correlated with the original

variable and uncorrelated with the disturbance part. This technique leads to consistent estimators.

In this thesis, the idea of both methods, new weightings and instrumental variable (grouping method as a special case) will be implemented to propose a new estimation method for fitting the multiple MEM with better estimates properties.

2.3 Existing approaches to fitting MEM

One of the most important aims while fitting the MEM is how to obtain a consistent estimator of model coefficients. To achieve these aims, numerous authors have explored various estimation methods to solve the problems such as the attenuation bias, identifiability, ill-posed and saddle point problems that may occurs in the relationship between two variables when both variables are subject to errors.

Several estimation methods are utilized to obtain consistent estimates in MEM, including but not limited to the instrumental variable estimation method (Madansky, 1959; Arellano & Bond, 1991), robust estimation methods (Theil, 1950; Siegel 1982), ordinary least squares (Fuller, 1987; Thompson & Carter, 2007), the method of moments (Gillard & Iles; 2005;), maximum likelihood estimation method (Carroll et al., 2007; Kendell & Stuart, 1979; Cheng & van ness, 1999), generalized maximum entropy (Al-Nasser, 2003), and the Bayesian Nonparametric Learning framework that is robust to mismeasured covariates (Charita & Damoulas, 2023). Additionally, recent advancements include the Weighted Exponential Distribution (EB-WE) and the Renyi Entropy-Based Weighted Exponential Distribution (REB-WED) proposed by Al-Zoubaidi and Al-Nasser (2024).

2.3.1 Instrumental variable estimation

The instrumental variable (IV) estimation method provides a consistent MEM estimate. To ensure consistency in estimating the model parameters, this method seek a set of IV that are highly correlated with the explanatory variables but uncorrelated with the error terms. Consider the model given in Equation (2.2), and assume that $\{Z_j, j = 1, 2, \dots, k\}$ are k independent IV each with n observations, such that, these IV are asymptotically correlated with X and uncorrelated with u.

Therefore, we have

$$\begin{aligned} plim \left(\frac{1}{n} Z'X \right) &= \Sigma_{ZX} \\ plim \left(\frac{1}{n} Z'u \right) &= 0 \end{aligned}$$

Then the consistent IV estimator of the slope parameter in the simple MEM will be:

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y$$

A variable that is likely to satisfy the two requirements of an IV is the Wald-Type estimators including the two-groups and the three groups' estimations methods. Wald (1940) introduced the two-group method to extract additional information about the precision ratio. He selects the IV Z to be in following form.

$$Z = \begin{cases} 1 & ; \text{if } x_j \geq \text{Median} \{x_j; j = 1, 2, \dots, n\} \\ -1 & ; \text{if } x_j < \text{Median} \{x_j; j = 1, 2, \dots, n\} \end{cases}$$

Based on this selection, then the two groups' estimators will be:

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y = \begin{bmatrix} n & n\bar{X} \\ 0 & \frac{n}{2}(\bar{X}_2 - \bar{X}_1) \end{bmatrix}^{-1} \begin{bmatrix} n\bar{Y} \\ \frac{n}{2}(\bar{Y}_2 - \bar{Y}_1) \end{bmatrix}$$

which can be written as

$$\widehat{\beta}_{IV} = \begin{bmatrix} \bar{Y} - \left(\frac{\bar{Y}_2 - \bar{Y}_1}{\bar{X}_2 - \bar{X}_1} \right) \bar{X} \\ \left(\frac{\bar{Y}_2 - \bar{Y}_1}{\bar{X}_2 - \bar{X}_1} \right) \end{bmatrix}$$

where (\bar{X}, \bar{Y}) is the sample means of the explanatory and response variables, respectively. (\bar{X}_1, \bar{Y}_1) and (\bar{X}_2, \bar{Y}_2) are the sample means of the lower and upper groups, respectively. This will be a consistent estimator of the MEM parameters.

Nair and Shrivastava (1942) expanded upon this by proposing a generalized grouping method, which involves plotting the means of the first and last thirds of the observations to achieve a more accurate slope estimate than Wald's method. In this case the IV is:

$$Z = \begin{cases} 1 & ; \text{if } x_j \geq \{x_j: j = 2\frac{n}{3} + 1, \dots, n\} \\ 0 & ; \text{if } x_j < \{x_j: j = n/3 + 1, \dots, 2n/3\} \\ -1 & ; \text{if } x_j < \{x_j: j = 1, 2, \dots, n/3\} \end{cases}$$

The weighted Wald estimation method will be the main estimation method that will be used in this thesis. Further details and extensions will be discussed in the next chapter. The advantage of these methods over the other methods is the simplicity in computation and the non-dependent of the normality assumption. However, this method may not give optimal solutions if there are possible outliers in the data. A series of innovative similar approaches have been introduced to advance the estimation of parameters in MEM. Al-Nasser et al. (2005) developed a robust Wald-type estimator for the MEM, drawing upon the concept of L-statistics by applying trimmed and Winsorized means. The method suggested the transformation statistics to minimize the impact of the extreme values in the data, avoiding spurious outliers. In a

closely related effort, Al-Nasser and Al-Haj Ebrahim (2005) introduced a novel AM estimator specifically designed for MEM, which was further enhanced by Ghapor et al. (2015) by the proposed and refined version of the AM procedure to enhance the estimation of the slope parameter in functional MEM. Furthermore, Al-Radaideh et al. (2010) introduced an innovative sampling technique known as L Ranked Set Sampling (LRSS), which was employed to estimate the MEM parameters using Wald-type estimators, further contributing to the robustness and efficiency of parameter estimation in such models. Also, Al-Nasser (2012) proposed a robust grouping method leveraging the maximum entropy median for enhanced estimation accuracy in his study. Schennach and Hu (2013) highlighted the identifiability of fully nonparametric classical errors-in-variables models from regressor and dependent variable data alone, barring specific parametric forms. This was followed by a proposal of a two-stage information theoretic approach for modelling relationships between two error-prone variables, with or without equation error, in Al-Nasser (2014).

2.3.2 Robust Theil-Type Estimators

Theil (1950) suggested the repeated median method for estimating MEM, which is notable for its simplicity and robustness compared to classical estimates. This method can be described to be as follows:

Step 1 Find all possible pairs of observations, assuming that all x_i 's are distinct, and

$$\text{given by } B_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad i, j = 1, 2, \dots, n \quad i < j, \quad \text{which yields } \binom{n}{2}$$

slope values.

Step 2 Find the median of the slopes that were obtained in (a), i.e.,

$$\hat{\beta} = \text{med}(B_{ij})$$

the intercept can be computed from

$$\hat{\alpha} = med(y_i - \hat{\beta}x_i)$$

However, Maritz (1979) extends Theil's ideas to propose an estimate of the intercept term by taking the median of all pairwise intercept estimates, i.e.,

$$A_{ij} = \frac{(x_i y_j - x_j y_i)}{x_i - x_j}, \quad i, j = 1, 2, \dots, n \quad i < j$$

hence

$$\hat{\alpha} = med(A_{ij})$$

Meanwhile, Siegel (1982) has applied Theil's method to estimate values using a repeated median approach. This method can be described to be as follows:

Step 1 Find all possible pairs of all observations, as given in step.1 of Theil.

Step 2 Find the median of the cross medians by

$$\hat{\beta}_i = med(B_{ij}) \quad i = 1, 2, \dots, n-1.$$

Step 3 Determine the estimate of the slope by taking

$$\hat{\beta} = med(\hat{\beta}_i).$$

The intercept can be estimated by using the same formula

2.3.3 Maximum Likelihood Estimation method

The maximum likelihood estimation (MLE) stands out as a widely accepted approach, endorsed by numerous researchers. Lindley (1947) first demonstrated the inability to obtain maximum likelihood estimators (MLE) due to the fact that these parameters are unidentified and remarked on the need to make some additional assumption(s) about the parameter values to alleviate this difficulty. Extensive

bibliographies that clarified this problem were given by Madansky (1959), Kendall and Stuart (1979), Cheng and Ness (1999), and Fuller (1987).

The later state that there are six side assumptions found in the literature, any one of which will make the normal structural no-equation-error model is identifiable. The most common assumption is that the ratio of the residual variances $\lambda = \sigma_1^2/\sigma_2^2$ is known. Other assumptions are the reliability ratio is known, either σ_1^2 , σ_2^2 or both are known and the last prior assumption that could be added is that the intercept is known.

However, for the equation-error model, the new formulation of the main equation by adding the error term might mean that at least some of the side assumptions no longer imply the identify ably of the slope. In fact, among the six side assumptions that can be used in the no-equation-error model, $\lambda = \sigma_1^2/\sigma_2^2$ and σ_1^2 are not enough to be known in order to identify the slope. More details can be found in Chen and Van Ness (1999). The MLE solutions for the normal structural MEM under different assumptions are given in Table 2.1. Therein we use the quantities

$$\bar{x} = n^{-1} \sum_{i=1}^n x_i \quad , \quad \bar{y} = n^{-1} \sum_{i=1}^n y_i$$