Pakistan Journal of Statistics and Operation Research

Weighted Grouping Estimation Method for Fitting Multiple Structural Regression Model when all Variables are Subject to Errors

Ro'ya Al Dibi'i[1], Rosmanjawati Abdul Rahman[2*], Amjad Al-Nasser[3]

* Corresponding Author

[1] School of Mathematical Sciences, USM, Penang, Malaysia, Royaaldibii@student.usm.my
[2] School of Mathematical Sciences, USM, Penang, Malaysia, rosmanjawati@usm.my
[3] Department of Statistics, Yarmouk University, Jordan, amjadn@yu.edu.jo

The Measurement Error Model (MEM) is employed to fit the relationship between two or more variables when all variables are subject to measurement errors. In the specific case of only two variables, this model is referred to as the Error in Variables model. This paper proposes two new estimation methods for a multiple structural measurement error model, applicable when all variables are subject to errors. The proposed methods, the Repetitive Weighted Grouping and the Iterative Weighted Grouping, are extensions of the Wald estimation method. To evaluate the performance of these new estimators compared to classical estimators-namely, the Maximum Likelihood Estimator (MLE) and the Method of Moments (MOM), a Monte Carlo experiment was conducted. The simulation results showed that the proposed estimators outperform the classical estimators in terms of root mean square error and bias. Additionally, real data analysis was performed to assess the relationships between national GDP, unemployment rate, and human development index using the proposed estimation methods. The results reveal that, based on mean square error (MSE), the proposed methods with r =3 and r =4 yield more accurate estimators than other methods in weight case 1, while the proposed method with r =4 proves more accurate in weight case 2. Furthermore, the proposed procedures demonstrate greater efficient than MLE and MOM in fitting the model.

Key Words: Measurement Error Model, Weighted Grouping Method, Wald Estimators, Iterative Estimator, Unemployment Rate, Monte Carlo Simulation.

1. Introduction

When modelling the relationship between two variables, the structural Measurement Error Model (MEM) (Adusumilli and Otsu, (2018); Fuller, (1987)) can be used as an extension of the standard linear regression model by accounting for independent errors in both the response and predictor variables. This study explores linear MEMs with vector-valued explanatory variables, i.e., models involving more than one x variable, by employing new estimation methods- the Repetitive Weighted Grouping and the Iterative Weighted Grouping, which are extensions of the Wald estimation method. These methods represent a modification of the classical MEM model.

Consider the following equation error model:

$$\eta_i = \alpha + \beta_{i1}\xi_{i1} + \beta_{i2}\xi_{i2} + \cdots + \beta_{ik}\xi_{ik} \qquad i = 1,2,\dots,n \quad , \; j = 1,2,\dots,k \tag{1}$$

where

$$y_i = \eta_i + \epsilon_i \quad \text{and} \quad x_{ij} = \xi_{ij} + \delta_{ij} \quad i = 1,2,\dots,n \quad , \; j = 1,2,\dots,k \tag{2}$$

The measurement errors $(\delta_{ij}, \epsilon_i)$ are random vectors that are independent and identically distributed, while the latent variable ξ_{ij} is generally assumed to be independent and normally distributed. However, the true distribution of ξ_{ij}

may deviate from normality in cases of skewness, outliers, or multimodality. Therefore, using more flexible models as alternatives to the standard one can be advantageous (Cabral et al., (2020)).

The main issue in (1) is estimating the unknown parameters α and β. (Cao et al., (2018b)) provides a comprehensive review of this issue. Unless further prior information is taken into consideration, standard estimation methods like Maximum Likelihood Estimation (MLE) fail to accurately estimate the unknown parameters of MEM. Consequently, researchers have been exploring alternative estimation methods to overcome the difficulties associated with MLE (Cheng and Van Ness, (1999); Adusumilli and Otsu, (2018); Ahmad and Ahmad, (2019); Salem, (2018); Fuller, 1987)). In this paper, an iterative estimation and weighted grouping method are proposed to fit various structural MEMs.

Measurement error, which occurs when an important variable is not correctly observed, is a significant issue that frequently raises concerns about the validity of an analysis. While there are numerous methods to address the effects of measurement inaccuracy, these methods lose their validity when their fundamental assumptions are violated. The problem is exacerbated when assumptions, such as those pertaining to the distribution of error terms, are difficult or impossible to assess using the available data. In the measurement error literature, an additive model with normally distributed errors is typically assumed. Although this assumption is simple and appealing, it is often incorrect in practical applications (Spicker et al., (2021)).

Numerous authors have discussed various techniques for estimating structural MEMs. The least squares and MLE methods are among the most popular. The MLE method is applied (by Schennach and Hu, (2013)) after certain assumptions are considered. Madansky, (1959) also provided a detailed explanation of how to use MLE to fit a straight-line model when both variables are measured with errors. Adusumilli and Otsu, (2018) offered a broad overview of the normal theory for structural MEMs. Additionally, Al-Nasser, (2004) demonstrated that the General Maximum Entropy (GME) method outperforms the Partial Least Squares (PLS) approach in terms of mean squares of error (MSE) when analysing distributions without relying on conventional assumptions.

Nonparametric or semiparametric approaches are important because they provide flexible strategies to mitigate the impacts of measurement error by avoiding assumptions about the distribution of the error components. Green, (2011) highlighted that adapting to non-normal errors is a significant area of research. Contributors to this field include Yi, (2017), Schennach, (2013), and Xu et al., (2017). On the other hand, Cao et al., (2018a) proposed a discretionary Expectation-Maximization (EM) algorithm combined with an empirical Bayesian technique to generate maximum likelihood estimates for MEMs with or without equation error. Similar results were achieved (by Cao et al., (2018b)), who developed an iterative maximum likelihood estimation procedure using the EM algorithm for heteroscedastic MEMs. An alternative estimation method, suggested (by Nair and Shrivastava, (1942)), involves generalized average grouping by plotting the first-third and last-third means of all observations, providing a more accurate slope estimate than Wald's method. Recently, information theory has been applied (by Al-Nasser, (2004, 2005)). Other authors (by Al-Nasser, (2011); (Carroll et al., 2007)) have proposed non-parametric estimators of regression functions using data contaminated by measurement errors. Furthermore, robust non-parametric estimation procedures were suggested (by Al-Nasser et al., (2016); Al-Nasser, (2012); Xu et al., (2017)). Additional insights into various estimation techniques within the context of MEMs have been provided (by Surajit, (2015); Gillard, (2010); Carroll et al., (2006); Adusumilli and Otsu, (2018); Bartlett, (1949); Wiedermann et al., (2018); Green, (2011)).

This article proposes two new non-parametric estimation methods: the repetitive weighted grouping and the iterative weighted grouping. The performance of these methods is demonstrated through Monte Carlo simulations and real data applications in estimating the parameters of multiple linear regression model with two independent variables in the presence of measurement errors. The rest of the paper is organized as follows: Section 2 provides a review of the Maximum Likelihood Estimation (MLE) and Method of Moments (MOM) approaches. Section 3 details the newly proposed procedures, including the iterative and weighted grouping (weighted Wald-type). Section 4 showcases the performance of these methods through Monte Carlo simulations, followed by a real data application in Section 5. Finally, Section 6 concludes the article.

2. Classical Estimation Methods for Multiple MEM

This section briefly discusses the common estimation techniques used for fitting multiple structural measurement error models: The Maximum Likelihood Estimation (MLE) and Method of Moment (MOM).

2.1 Maximum Likelihood Estimator

MLE is the most commonly used estimation method. The estimation problem for the simple MEM using the MLE was initially addressed (by Lindley, (1947)) and further elaborated (by Kendall and Stuart, (1979)). A major drawback of this method is that it requires the assumption of normality for the unknown error terms. Consider the models given in (1) and (2) written in matrix form:

$$\boldsymbol{\eta} = \xi' \boldsymbol{\beta}; \;\; \boldsymbol{Z_t} = \boldsymbol{z_t} + \boldsymbol{\varepsilon_t} \tag{3}$$

where

$$\boldsymbol{Z_t} = (\boldsymbol{X_t}, \boldsymbol{Y_t}); \; \boldsymbol{\varepsilon_t} = (\boldsymbol{\delta_t}, \boldsymbol{\varepsilon_t}).$$

$\boldsymbol{\eta}$: Linear predictor or the expected value of the response variable.

$\boldsymbol{\xi}$**:** Vector of explanatory variables or predictors.

$\boldsymbol{\beta}$**:** Vector of coefficients.

$\boldsymbol{Z_t}$**:** Observed data vector at time t.

$\boldsymbol{z_t}$**:** Deterministic component or the part of the model that the predictors explained.

$\boldsymbol{\varepsilon_t}$**:** Error term at time t.

$\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ **:** The components of the observed data vector $\boldsymbol{Z_t}$**.**

$\boldsymbol{\delta_t}$ and $\boldsymbol{\varepsilon_t}$ **:** The components of the error vector $\boldsymbol{\varepsilon_t}$**.**

Then, under the multivariate normal distribution (i.e., $\boldsymbol{\varepsilon_t \sim NI(0, \Sigma_{\varepsilon\varepsilon})}$) assumption, the variance-covariance is known and given as:

$$\boldsymbol{\Sigma_{\varepsilon\varepsilon}} = \boldsymbol{\Upsilon_{\varepsilon\varepsilon}}\sigma^2$$

where

$\boldsymbol{\Sigma_{\varepsilon\varepsilon}}$**:** The covariance matrix of the error terms $\boldsymbol{\varepsilon_t}$**.**

$\boldsymbol{\sigma^2}$: Variance of the error terms.

$\boldsymbol{\Upsilon_{\varepsilon\varepsilon}}$**:** Known matrix representing the structure of the error covariance.

Accordingly, the unknown parameters can be estimated by finding the first derivative of the log-likelihood function of a random sample of size n:

$$L(\beta, \lambda, \theta;\, y, Z) \;=\; \prod_{i=1}^{n} \int f(y_i|X_i, U_i;\, \beta) f\left(Z_{i1}, \dots\dots, Z_{ir_i}|X_i, U_i;\, \lambda\right) f(X_i|\, U_i;\, \theta)\, dx_i$$

$$\boldsymbol{Log\ L = c - \frac{n}{2}\ log|2\pi \Upsilon_{\varepsilon\varepsilon}\sigma^2| - \frac{1}{2\,\sigma^2} \sum_{t=1}^{n} (Z_t - z_t)\, \Upsilon_{\varepsilon\varepsilon}^{-1} (Z_t - z_t)'}.$$

Solving the first order conditions of the log likelihood function, the unknown parameters can be estimated as:

$$\boldsymbol{\hat{\beta}} = \left[\boldsymbol{M_{XX}} - (\boldsymbol{\hat{\lambda}} - \frac{\boldsymbol{1}}{\boldsymbol{n}}\ \boldsymbol{\alpha})\boldsymbol{S_{\delta\delta}}\right]^{\boldsymbol{-1}} \left[\boldsymbol{M_{XY}} - (\boldsymbol{\hat{\lambda}} - \frac{\boldsymbol{1}}{\boldsymbol{n}}\ \boldsymbol{\alpha})\boldsymbol{S_{\delta\epsilon}}\right]$$

where

$\boldsymbol{M_{zz}} = \frac{\mathbf{1}}{\boldsymbol{n-1}} (\boldsymbol{Z_t} - \overline{\boldsymbol{Z}})' (\boldsymbol{Z_t} - \overline{\boldsymbol{Z}})$ **;** $\hat{\lambda}$ is the smallest root of $|M_{XX} - \lambda S_{\varepsilon\varepsilon}|$=0; and $S_{\varepsilon\varepsilon}$ is an unbiased estimator of $\boldsymbol{\Sigma_{\varepsilon\varepsilon}}$.

2.2 Method of Moments for Multiple MEM

The method of moments (MOM) is a commonly used technique for estimating MEM. Originally introduced (by Geary, (1942)), MOM relied on sample and population moments, and was later modified to incorporate cumulates in his research. Although Drio, (1951) also applied MOM in his study, it has not been frequently cited. In more recent years, researchers such as Cragg, (1997), Gillard and Iles, (2006), Van Montfort, (1988), and Pal, (1980) have explored the use of moments to develop optimal estimators, particularly those based on higher moments. Dunn, (2000) developed several slope estimators using the MOM but did not provide details on estimators based on higher moments. Following Pal, (1980), the MOM estimator for the model given in (1) can be derived by computing the deviations of all variables specified in the model:

$$\boldsymbol{\eta' = \eta - \overline{\eta}; \xi'_i = \xi_i - \overline{\xi_\iota}; y' = y - \overline{y}} \textbf{ and } \boldsymbol{x'_i = x_i - \overline{x_\iota}} \quad (4)$$

Also, the error terms are assumed symmetrically distributed, then

$$\boldsymbol{E\left(y'x_i'^2\right) = \sum_{j=1}^{k} \beta_j E\left(x_j' x_i'^2\right); i = 1, 2, \dots, m.}$$

Therefore,

$$\boldsymbol{A\beta = B.}$$

Hence

$$\boldsymbol{\beta = A^{-1}B}$$

where

$$\boldsymbol{A = \left(\left(a_{ij}\right)\right) \; with \; a_{ij} = E\left(x_i'^2 x_j'\right)}$$

$$\boldsymbol{B' = \left(E\left(y'x_1'^2\right), E\left(y'x_2'^2\right), \dots, E\left(y'x_m'^2\right)\right)}$$

and

$$\boldsymbol{\beta' = (\beta_1, \beta_2, \dots, \beta_m).}$$

Finally, $\hat{\beta} = \hat{A}^{-1}\hat{B}$ a consistent estimator provided $|A| \neq 0$; where $\hat{A} \; and \; \hat{B}$ are the sample estimates of A and B. Therefore, unless additional information about the relationship beyond the observations is available, only MOM estimators can be used, and the variances of such estimators remain unknown.

3. The New Proposed Procedures.

This article proposes two new procedures for fitting a multiple structural measurement error model when all variables are subject to errors. These new estimation methods extend the Wald estimation method, commonly referred to as the grouping method which involves splitting the data into two or three groups and estimating the slope of the MEM based on the group centers (by Wald, (1940), Gillard, (2010)). The proposed methods include a weighted grouping technique and an iterative weighted procedure.

3.1 The Weighted Grouping Method

The weighted grouping method is utilized to avoid the unpredictability of individual results and multicollinearity across explanatory variables by multiplying the weight of each group by its sample mean in the Wald equation and iterative equation. The general idea behind this process is as follows:

- Sort the data from smallest to largest with their corresponding values $y_i's,\ i = 1,2,\dots,n$.
- Split the data into r-subgroups of similar size (i.e. the sub-sample size is k) so that $r \le \left[\frac{n}{2}\right]$.
- Compute the parameters, $\hat{\beta}_i$ as:

$$\hat{\beta}_i = \frac{w_{im}\bar{y}_{im} - w_{i(m-1)}\bar{y}_{i(m-1)}}{w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}}, \quad m = 1,2,\dots\dots,r,\ i = 1,2,\dots,n \tag{5}$$

where w_{im} is the weighted group.

- Determine the weight w_{im} for two cases:

(i) Case One: compute weight as:

$$\boldsymbol{w_{i(m-1)} = cov\left(x_{i(m-1)}, y_{(m-1)}\right)} \tag{6}$$

$$\boldsymbol{w_{im} = cov(x_{im}, y_m)} \tag{7}$$

(ii) Case Two: compute weight as:

$$w_{im} = \frac{\sigma^2_{x_{im}}}{\sum_{i=1}^{k} \sigma^2_{x_i}} \tag{8}$$

$$w_{i(m-1)} = 1 - w_{im} \tag{9}$$

where:

$$\sum(\boldsymbol{w_{im}} + \boldsymbol{w_{i(m-1)}}) = \mathbf{1}$$

Theorem: Assuming the model in (1) and (2) holds, the estimator based on Weighted Grouping Method given in (5) is unbiased estimator if and only if $w_{im} = w_{i(m-1)}$ in the first case, and $w_{im} = w_{i(m-1)} = 0.5$ in the second case (refer to Gupta and Amanullah, (1970)).

Proof:

$$\hat{\beta}_i = \frac{w_{im}\bar{y}_{im} - w_{i(m-1)}\bar{y}_{i(m-1)}}{w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}}$$

$$\hat{\alpha} = \bar{y} - \sum_{i=1}^{k} \hat{\beta}_i\, \bar{x}_i$$

$$E\left(\hat{\beta}_i\right) = E\left(\frac{w_{im}\bar{y}_{im} - w_{i(m-1)}\bar{y}_{i(m-1)}}{w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}}\right) \tag{10}$$

However,

$$E(\epsilon_i) = 0, \qquad E\left(\epsilon_i\epsilon_j\right) = 0,\ and \quad E(\epsilon_i^2) = \sigma^2.$$

Then,

$$\hat{\beta}_i = \beta_i + \frac{w_{im}\bar{\epsilon}_{im} - w_{i(m-1)}\bar{\epsilon}_{i(m-1)}}{w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}}$$

where

$\bar{\epsilon}_{im}$ and $\bar{\epsilon}_{i(m-1)}$ are the means of n_{im} and $n_{i(m-1)}$ disturbances ϵ_i.

By the theorem, when $w_{im} = w_{i(m-1)}$, and using the results (from Raj and Koerts, (1992)), we obtain:

 DOI: http://dx.doi.org/10.18187/pjsor.v20i3.4483

$$E\left(\hat{\beta}_i\right) = \beta_i ,$$

with associated variance given as:

$$Var\left(\hat{\beta}_i\right) = \frac{1}{\left(w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}\right)^2}\left\{Var\left(w_{im}\bar{y}_{im}, w_{i(m-1)}\bar{y}_{i(m-1)}\right)\right\}$$

$$= \frac{w_{im}{}^2 var(\bar{y}_{im}) + w_{i(m-1)}{}^2 var\left(\bar{y}_{i(m-1)}\right) - 2w_{im}\, w_{i(m-1)}\, cov\left(\bar{y}_{im}, \bar{y}_{i(m-1)}\right)}{\left(w_{im}\bar{x}_{im} - w_{i(m-1)}\bar{x}_{i(m-1)}\right)^2}. \quad (11)$$

Also,

$$[\, E\,(\hat{\alpha}) = E\left(\bar{y} - \sum_{i=1}^{k} \hat{\beta}_i\,\bar{x}_i\right)$$

$$= \left(\alpha + \sum_{i=1}^{k} \hat{\beta}_i\,\bar{x}_i - \sum_{i=1}^{k} \hat{\beta}_i\,\bar{x}_i\right) = \alpha\ ] \quad (12)$$

with

$$var\,(\hat{\alpha}) = var\left(\bar{y} - \sum_{i=1}^{k} \hat{\beta}_i\,\bar{x}_i\right)$$

$$= var(\bar{y}) + \bar{x}_i\, var\left(\hat{\beta}_i\right) - 2\,cov\,\left(\bar{y}_{im}, \bar{y}_{i(m-1)}\right) \quad (13)$$

3.2 The Iterative Weighted Procedure

Meanwhile, the proposed iterative weighted procedure is an extension of the repetitive estimation method described (by Al-Dibi'i and Al-Nasser, (2019)). The general idea of this procedure can be summarized as follows:

- Sort the y values in ascending order, from smallest to largest, along with their associated $\left(x_{1[i]}, x_{2[i]}\right)$ values, where $i = 1, 2 \ldots, n$.
- Divide the data into r-subgroups of equal size (i.e the sub-sample size is k) such that $r \leq \left[\frac{n}{2}\right]$.
- Compute the mean for each subgroup $\left(\bar{x}_{1j}, \bar{x}_{2j}, \bar{y}_j\right); j = 1, 2, \ldots, r$.
- Estimate the unknown parameters. At this stage, this article proposes two procedures, which are described in the following subsections.

3.2.1 Repetitive Weighted Grouping Procedure

The idea of this procedure is to take the center of all possible slopes (Figure 1), as has been suggested (by Al-Dibi'i and Al-Nasser, (2019)). This can be done by defining the j-th slope as:

$$\hat{\beta}_{kj} = \frac{w_{mj}\bar{y}_{mj} - w_{(m-1)j}\bar{y}_{(m-1)j}}{w_{mj}\bar{x}_{mj} - w_{(m-1)j}\bar{x}_{(m-1)j}}, \quad j = 1, 2, \ldots \ldots n, \quad m = 1, 2, \ldots \ldots, r, \quad k = 1, 2, \ldots, \binom{r}{2}$$

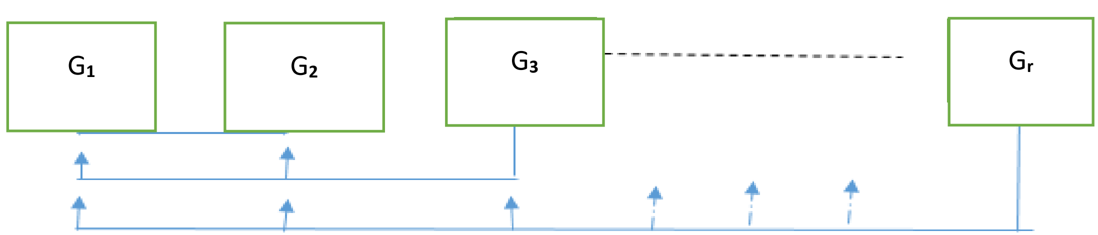

Figure 1: All possible slopes between the subgroups

- Finally, the unknown parameters of MEM can be estimated as:

$$\hat{\beta}_{kj} = \frac{1}{\binom{r}{2}} \sum_{j=1}^{\binom{r}{2}} \hat{\beta}_{\boldsymbol{kj}} \quad \text{and} \quad \hat{\alpha} = \bar{y} - \sum_{k=1}^{r-1} \hat{\beta}_k \, \bar{x}_k \tag{14}$$

3.1.2 An Iterative Weighted Grouping Procedure

The idea in this procedure is to compute the pairwise slopes continuously and gradually from each sub-group to another sub-group as illustrated in Figure 2. Therefore, the j-th slope iteratively as:

$$\hat{\beta}_{kj} = \frac{w_{mj}\bar{y}_{mj} - w_{(m-1)j}\bar{y}_{(m-1)j}}{w_{mj}\bar{x}_{mj} - w_{(m-1)j}\bar{x}_{(m-1)j}}, \quad j = 1, 2, \dots \dots n, \quad m = 1, 2, \dots \dots, r, \quad k = 1, 2, \dots, (r-1) \tag{15}$$

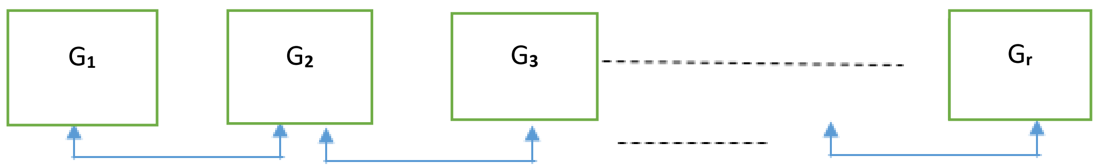

Figure 2: Pairwise slope between the subgroups

- Finally, the unknown parameters of MEM can be estimated as:

$$\hat{\beta}_{kj} = \frac{1}{(r-1)} \sum_{j=1}^{(r-1)} \hat{\beta}_{\boldsymbol{kj}} \quad \text{and} \quad \hat{\alpha} = \bar{y} - \sum_{k=1}^{r-1} \hat{\beta}_k \, \bar{x}_k \tag{16}$$

4. Monte Carlo Experiment

Two random samples, consisting of inlier and outlier samples, were generated using Python, each based on 10,000 iterations with a sample size of n, from the standard normal MEM described in (1). These samples were analyzed under the following procedures and assumptions.

- Order the data from smallest to largest with their respective associated Yi values, $i = 1, 2, \dots, n$.

  where:

$$\eta_i = \alpha + \beta_{i1}\xi_{i1} + \beta_{i2}\xi_{i2} \qquad i = 1, 2, \dots, n$$

$$y_i = \eta_i + \epsilon_i$$

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

where

$$x_{i1} = \xi_{i1} + \delta_{i1}$$

$$x_{i2} = \xi_{i2} + \delta_{i2}$$

- Set the initial values as $\alpha = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\sigma_\epsilon^2 = 1$, $\sigma_{\delta 1}^2 = 1$ and $\sigma_{\delta 2}^2 = 1$.
- Generate the error terms from a standard normal distribution.
- Consider three different data sizes: $n = 100$, 200 and 500.
- Contaminate the data with outliers. At each step a certain percentage (10%) of the observations were deleted and replaced with outliers generated according to the following different cases:

(i) Outliers only in y ($\epsilon_i \sim N(0, \sigma_\epsilon^2)$, $\sigma_\epsilon^2 = 16$.
(ii) Outliers only in x_1 ($\delta_1 \sim N(0, \sigma_{\delta 1}^2)$, $\sigma_{\delta 1}^2 = 16$.
(iii) Outliers only in x_2 ($\delta_2 \sim N(0, \sigma_{\delta 2}^2)$, $\sigma_{\delta 2}^2 = 16$.
(iv) Outliers in both $x_1 and\ x_2$ ($\delta_1 \sim N(0, \sigma_{\delta 2}^2)$ and ($\delta_2 \sim N(0, \sigma_{\delta 2}^2)$, $(\sigma_{\delta 1}^2, \sigma_{\delta 2}^2)$ = (16, 16).
(v) Outliers in y, $x_1 and\ x_2$, $(\sigma_\epsilon^2, \sigma_{\delta 1}^2, \sigma_{\delta 2}^2)$ = (16, 16, 16).

The properties of these estimators were investigated using the simulated bias and mean square error, defined as

$$Bias = \frac{1}{10000}\sum_{i=1}^{10000}(\hat{\mu}_i - \mu); \qquad MSE = \frac{1}{10000}\sum_{i=1}^{10000}(\hat{\mu}_i - \mu)^2 \tag{17}$$

where $\hat{\mu}_i$ is the estimates given by one of the proposed estimators for the i[th] sample.

Table 1-6 present the bias and MSE values of $\hat{\alpha}$ and $\hat{\beta}$ for each contaminated case across different sample sizes: n = 100, 200 and 500. The simulated results indicate that weighted technique 1 outperformed technique 2 when there were no outliers, while technique 2 showed better performance in the presence of outliers. Overall, the bias and mean squared error (MSE) of the estimations decreased as the sample size increased. Compared to classical methods like Maximum Likelihood Estimation (MLE) and Method of Moments (MOM), the newly proposed weighted grouping procedures exhibited lower bias and MSE values. Additionally, the study found that weighted technique 2 was more efficient than technique 1 in the presence of outliers.

Table 1: The Bias and MSE of $\hat{\alpha}$ and $\hat{\beta}$ for samples without outlier.

n	Parameter	Statistic	Weight case 1				Weight case 2				Classical	
			Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	MLE	MOM
100	$\hat{\alpha}$	Bias	0.0001	0.0001	0.0005	0.0001	-0.0088	-0.0045	0.0023	-0.001	-0.0014	-0.0495
		MSE	0.0035	0.0026	0.0207	0.013	0.0820	0.0839	0.0518	0.0219	0.2334	0.3387
	$\hat{\beta}_1$	Bias	-0.0186	-0.0198	0.0339	-0.02	-0.0057	0.0136	0.0488	-0.0194	0.327	0.6429
		MSE	0.0737	0.0389	0.201	0.1245	0.0623	0.0512	0.0279	0.0381	0.0782	0.5076
	$\hat{\beta}_2$	Bias	-0.0295	-0.0295	0.008	-0.0296	0.0622	-0.0459	0.0066	-0.0296	-0.8378	0.2111
		MSE	0.0873	0.0873	0.0915	0.0877	0.0592	0.0316	0.0209	0.0219	0.8187	0.6803
200	$\hat{\alpha}$	Bias	-0.0001	0.0001	-0.032	0.0024	0.0022	0.0001	-0.0028	0.0021	0.0049	0.0218
		MSE	0.0007	0.0007	0.0092	0.0055	0.0212	0.0264	0.0258	0.0356	0.0926	0.3036
	$\hat{\beta}_1$	Bias	-0.0098	-0.0095	0.0188	-0.0078	0.002	-0.0005	-0.0017	-0.0098	0.1817	0.3847
		MSE	0.0193	0.0190	0.0707	0.0784	0.0292	0.0104	0.0206	0.0247	0.7759	0.4736
	$\hat{\beta}_2$	Bias	-0.0149	-0.0140	0.0037	-0.0131	0.0249	0.0093	0.0028	-0.0159	-0.8608	0.0768
		MSE	0.0443	0.0440	0.0901	0.0743	0.0499	0.0185	0.0198	0.0245	0.7925	0.5327
500	$\hat{\alpha}$	Bias	0.00001	0.0001	0.051	-0.0014	0.0001	0.0001	-0.001	0.0006	0.0009	0.0034
		MSE	0.0001	0.0001	0.0083	0.0028	0.0041	0.0112	0.0109	0.0201	0.035	0.589
	$\hat{\beta}_1$	Bias	-0.004	-0.0034	0.0063	-0.0055	0.0006	0.0006	0.0067	0.0069	0.1203	0.9089
		MSE	0.0079	0.0078	0.0172	0.0069	0.0202	0.0102	0.0160	0.0184	0.3831	0.3516
	$\hat{\beta}_2$	Bias	-0.006	-0.0056	0.0026	-0.0013	0.0044	0.0048	0.0014	0.002	-0.8712	0.0332
		MSE	0.0179	0.0166	0.0811	0.0579	0.1518	0.0116	0.0096	0.0087	0.7791	0.424

Table 2: The Bias and MSE for $\hat{\alpha}$ and $\hat{\beta}$ when $\sigma_{\delta 1}^2 = 16$ with outliers in x_1.

n	Parameter	Statistic	Weight case 1	Weight case 2	Classical

			Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	MLE	MOM
100	$\hat{\alpha}$	Bias	-0.0001	0.0001	0.0002	0.0047	0.0009	-0.0034	0.0034	0.0005	-0.009	0.1412
		MSE	0.0048	0.0045	0.2862	0.2457	0.1755	0.1864	0.5892	0.5737	0.9498	0.5451
	$\hat{\beta}_1$	Bias	-0.0198	-0.0187	0.0326	-0.0195	0.0148	0.0045	0.0352	-0.0197	-0.8448	0.4498
		MSE	0.0378	0.0391	0.1179	0.1300	0.1448	0.1254	0.1461	0.1460	0.9722	0.7382
	$\hat{\beta}_2$	Bias	-0.0295	-0.0291	0.0148	-0.0296	0.0591	0.0452	0.0152	-0.0296	0.1294	0.1836
		MSE	0.0871	0.0870	0.2011	0.1989	0.1712	0.1663	0.1937	0.1901	0.3265	0.2918
200	$\hat{\alpha}$	Bias	-0.001	0.0001	0.0311	-0.0018	0.0004	-0.0005	0.0006	0.0058	0.055	0.0344
		MSE	0.0009	0.0022	0.2349	0.2050	0.1031	0.1004	0.4022	0.3127	0.349	0.2656
	$\hat{\beta}_1$	Bias	-0.0098	-0.0093	0.0166	-0.0099	0.0021	0.0011	0.0249	-0.0099	0.2975	0.3279
		MSE	0.0194	0.0191	0.1081	0.995	0.1342	0.1309	0.0841	0.0711	0.2775	0.583
	$\hat{\beta}_2$	Bias	-0.0149	0.0190	0.0042	-0.0149	0.0193	0.0044	0.0036	-0.0149	0.0004	0.0733
		MSE	0.0443	0.0440	0.1701	0.1544	0.1016	0.1498	0.0914	0.0901	0.2263	0.2811
500	$\hat{\alpha}$	Bias	-0.0001	-0.0002	0.0001	0.0021	0.0007	-0.0004	0.0001	0.0002	0.0147	-0.0567
		MSE	0.0002	0.0002	0.2209	0.1998	0.0224	0.0302	0.1292	0.1234	0.2814	0.2213
	$\hat{\beta}_1$	Bias	-0.004	-0.0033	0.0069	-0.004	0.0006	0.0006	0.0075	0.0069	0.2372	0.241
		MSE	0.0079	0.0066	0.0920	0.0900	0.0444	0.0873	0.0332	0.0311	0.2645	0.2955
	$\hat{\beta}_2$	Bias	-0.006	-0.005	0.0011	-0.006	0.0014	0.0064	0.0011	0.0013	-0.5327	0.0138
		MSE	0.0179	0.015	0.0804	0.0778	0.0283	0.0107	0.0687	0.0581	0.1875	0.1794

Table 3: The Bias and MSE for $\hat{\alpha}$ and $\hat{\beta}$ when $\sigma_{\delta 2}^2 = 16$ with outliers in x_2.

n	Parameter	Statistic	Weight case 1				Weight case 2				Classical	
			Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	MLE	MOM
100	$\hat{\alpha}$	Bias	0.001	0.0001	-0.0662	-0.0040	0.0045	0.0026	0.0043	-0.0079	-0.0099	-0.3700
		MSE	0.0073	0.0070	0.1185	0.0936	0.1502	0.1394	0.0714	0.0487	0.9089	0.6376
	$\hat{\beta}_1$	Bias	-0.0185	-0.0192	-0.0602	-0.0193	0.0075	0.0106	0.0174	-0.0193	0.1614	0.1955
		MSE	0.0588	0.0419	0.1209	0.1200	0.2963	0.1995	0.0415	0.0228	0.5679	0.7781
	$\hat{\beta}_2$	Bias	-0.0296	-0.0274	0.0059	-0.0296	0.0268	0.0498	0.011	-0.0296	-0.0974	-0.8605
		MSE	0.0874	0.0712	0.0148	0.0879	0.2785	0.2178	0.0432	0.0416	0.2595	0.4314
200	$\hat{\alpha}$	Bias	0.0001	-0.0002	-0.0078	-0.0007	0.0005	-0.0001	0.0005	-0.0014	-0.0019	-0.2851
		MSE	0.0013	0.00012	0.097	0.0026	0.0817	0.1255	0.0612	0.0324	0.8886	0.6183
	$\hat{\beta}_1$	Bias	-0.0099	-0.0098	-0.1023	-0.0098	0.0022	0.0028	0.0197	-0.0099	0.1596	0.1693
		MSE	0.0197	0.0193	0.096	0.0911	0.1924	0.1291	0.0883	0.0197	0.4712	0.6722
	$\hat{\beta}_2$	Bias	-0.0148	-0.0146	0.0032	-0.0149	0.0116	0.0217	0.0038	-0.0149	-0.0758	-0.6878
		MSE	0.0466	0.0358	0.0023	0.0025	0.2033	0.1463	0.0213	0.0404	0.2179	0.3466
500	$\hat{\alpha}$	Bias	-0.0001	-0.0001	0.001	-0.0015	0.007	-0.003	0.041	-0.0011	0.0775	-0.1193
		MSE	0.0001	0.0001	0.009	0.0076	0.0079	0.0035	0.0302	0.0301	0.5917	0.4216
	$\hat{\beta}_1$	Bias	-0.004	-0.0036	0.0078	-0.0085	0.0008	0.0007	0.0071	-0.004	0.2358	0.1679
		MSE	0.008	0.0073	0.0513	0.0144	0.0804	0.0103	0.0282	0.0178	0.2825	0.6641
		Bias	-0.006	-0.0054	0.0014	0.0026	0.0051	0.0038	0.0014	-0.006	-0.9269	-0.4754

$\hat{\beta}_2$	MSE	0.0181	0.0163	0.0019	0.0013	0.0905	0.0391	0.0201	0.0169	0.1333	0.2971

Table 4: The Bias and MSE for $\hat{\alpha}$ and $\hat{\beta}$ when $(\sigma^2_{\delta 1}, \sigma^2_{\delta 2})$ = (16, 16) with outliers in both (x_1, x_2).

n	**Parameter**	**Statistic**	**Weight case 1**				**Weight case 2**				**Classical**	
			Repetitive r = 3	**Repetitive r = 4**	**Iterative r = 3**	**Iterative r = 4**	**Repetitive r = 3**	**Repetitive r = 4**	**Iterative r = 3**	**Iterative r = 4**	**MLE**	**MOM**
100	$\hat{\alpha}$	**Bias**	-0.0003	0.0001	0.0071	-0.0075	0.0009	-0.0034	0.041	-0.346	-0.3903	-0.1765
		MSE	0.0099	0.0088	0.4007	0.3902	0.1755	0.1864	0.5009	0.4515	0.5573	0.4166
	$\hat{\beta}_1$	**Bias**	-0.0198	-0.0191	0.0615	-0.0195	0.0148	0.0045	0.022	-0.0197	-0.176	0.4669
		MSE	0.0392	0.0386	0.2053	0.2008	0.2448	0.2254	0.1478	0.1066	0.6702	0.5008
	$\hat{\beta}_2$	**Bias**	-0.0303	-0.0291	0.0088	-0.0299	0.0591	0.0452	0.2589	-0.0372	0.6054	-0.7889
		MSE	0.0972	0.0872	0.1908	0.1524	0.1712	0.1663	0.3124	0.1946	0.6821	0.4622
200	$\hat{\alpha}$	**Bias**	-0.0001	0.0002	0.0007	-0.0097	0.0004	-0.0005	-0.0022	-0.0052	-0.2256	-0.1529
		MSE	0.0015	0.0022	0.2862	0.2481	0.1031	0.1804	0.3808	0.2526	0.377	0.3547
	$\hat{\beta}_1$	**Bias**	-0.0098	-0.0015	0.0211	-0.0101	0.0021	0.0011	0.0244	-0.0099	-0.4022	0.3909
		MSE	0.0193	0.0190	0.1944	0.1085	0.1342	0.1309	0.0955	0.0923	0.208	0.4425
	$\hat{\beta}_2$	**Bias**	-0.0151	-0.012	0.0038	-0.0149	0.0193	0.0044	-0.003	-0.0149	0.1096	-0.4725
		MSE	0.0843	0.0818	0.1007	0.9444	0.1516	0.1498	0.0952	0.0944	0.5781	0.2472
500	$\hat{\alpha}$	**Bias**	-0.0004	-0.0002	0.001	-0.0019	0.0007	-0.0004	0.001	-0.0012	0.3113	-0.0946
		MSE	0.0002	0.0002	0.1992	0.1877	0.0924	0.1302	0.1933	0.1713	0.3353	0.2665
	$\hat{\beta}_1$	**Bias**	-0.004	-0.0033	0.0770	-0.046	0.0006	0.0006	0.0079	-0.004	0.118	0.1777
		MSE	0.0076	0.0066	0.1356	0.9979	0.0444	0.0873	0.0259	0.0184	0.173	0.2097
	$\hat{\beta}_2$	**Bias**	-0.006	-0.005	0.0026	-0.017	0.0014	0.0064	0.0014	-0.006	0.1554	-0.2105
		MSE	0.0181	0.0151	0.0622	0.0619	0.1283	0.1107	0.0728	0.0571	0.2385	0.1731

Table 5: The Bias and MSE for $\hat{\alpha}$ and $\hat{\beta}$ when $\sigma^2_{\epsilon} = 16$ with outliers in y.

n	**Parameter**	**Statistic**	**Weight case 1**				**Weight case 2**				**Classical**	
			Repetitive r = 3	**Repetitive r = 4**	**Iterative r = 3**	**Iterative r = 4**	**Repetitive r = 3**	**Repetitive r = 4**	**Iterative r = 3**	**Iterative r = 4**	**MLE**	**MOM**
100	$\hat{\alpha}$	**Bias**	0.0001	0.0001	0.0004	-0.0012	0.0045	0.0026	0.0806	-0.3095	-0.0566	-0.3533
		MSE	0.0029	0.0002	0.1921	0.1642	0.2502	0.2094	0.1321	0.1095	0.3392	0.8736
	$\hat{\beta}_1$	**Bias**	-0.0192	-0.0189	0.0447	-0.0195	0.0075	0.0106	0.0276	-0.0373	0.2062	0.2733
		MSE	0.0372	0.0350	0.258	0.1897	0.1963	0.1955	0.1709	0.1194	0.4143	0.3996
	$\hat{\beta}_2$	**Bias**	-0.0295	-0.0274	0.0095	-0.0296	0.0268	0.0498	0.0123	-0.0296	-0.3901	0.1234
		MSE	0.0872	0.0821	0.1606	0.1078	0.2485	0.2178	0.1004	0.1503	0.4639	0.2641
200	$\hat{\alpha}$	**Bias**	0.0003	-0.0001	0.0201	-0.0303	0.0005	-0.0001	-0.0004	-0.0073	-0.0235	0.0061
		MSE	0.0007	0.0001	0.0941	0.0761	0.2017	0.2055	0.0822	0.0788	0.1122	0.1229
		Bias	-0.0098	-0.0081	0.0201	-0.0099	0.0022	0.0028	0.02	-0.0103	0.0883	0.2431

	$\hat{\beta}_1$	MSE	0.0193	0.0176	0.123	0.1195	0.1924	0.1291	0.0723	0.0721	0.3886	0.2127
	$\hat{\beta}_2$	Bias	-0.0148	-0.0119	0.0042	-0.0149	0.0116	0.0217	0.0033	-0.0149	-0.330	0.0457
		MSE	0.0466	0.0358	0.0868	0.0844	0.2033	0.1463	0.0535	0.0445	0.2832	0.1022
500	$\hat{\alpha}$	Bias	-0.0001	-0.0001	0.011	0.0018	0.007	-0.003	0.001	0.0028	-0.0182	0.0872
		MSE	0.0001	0.0001	0.0781	0.0551	0.1979	0.1035	0.0383	0.0297	0.1081	0.1023
	$\hat{\beta}_1$	Bias	-0.004	-0.0036	0.0074	-0.008	0.0008	0.0007	0.0045	-0.006	0.0595	0.1686
		MSE	0.008	0.0078	0.0919	0.0910	0.1004	0.1103	0.0132	0.0119	0.1639	0.2108
	$\hat{\beta}_2$	Bias	-0.006	-0.0054	0.0015	-0.01	0.0051	0.0038	0.001	0.008	-0.2901	0.0365
		MSE	0.0181	0.0163	0.0212	0.0208	0.0905	0.0891	0.0106	0.0110	0.1553	0.0916

Table 6: The Bias and MSE for $\hat{\alpha}$ and $\hat{\beta}$ when $(\sigma^2_{\delta 1}, \sigma^2_{\delta 2}, \sigma^2_{\epsilon})$ = (16, 16, 16) with outliers in all (x_1, x_2, y).

n	Parameter	Statistic	Weight case 1				Weight case 2				Classical	
			Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	Repetitive r = 3	Repetitive r = 4	Iterative r = 3	Iterative r = 4	MLE	MOM
100	$\hat{\alpha}$	Bias	0.0002	0.0002	-0.0042	-0.0001	0.0044	0.0081	-0.0242	-0.0001	0.1915	0.3141
		MSE	0.0104	0.0107	0.4113	0.2068	0.2434	0.2299	0.5075	0.4072	0.8117	0.6594
	$\hat{\beta}_1$	Bias	-0.0188	-0.019	0.0465	-0.0194	0.0063	0.0033	-0.0532	-0.0192	-0.1953	0.7951
		MSE	0.0455	0.047	0.3683	0.2004	0.1392	0.1747	0.3336	0.1812	0.8682	0.6918
	$\hat{\beta}_2$	Bias	-0.0295	-0.0296	0.022	-0.0296	-0.0138	0.0839	0.0115	-0.0296	0.1169	-0.1741
		MSE	0.0872	0.0876	0.2462	0.1877	0.1272	0.1243	0.1262	0.1106	0.6714	0.2756
200	$\hat{\alpha}$	Bias	-0.0001	0.0002	0.0324	-0.013	-0.0002	-0.0018	0.0003	-0.0023	0.061	-0.3119
		MSE	0.0016	0.0015	0.2899	0.1921	0.1466	0.1361	0.4970	0.1569	0.7348	0.5576
	$\hat{\beta}_1$	Bias	-0.0098	-0.0098	0.0215	-0.0148	0.0023	0.0023	0.0243	-0.0099	0.1463	0.6997
		MSE	0.0198	0.0193	0.2798	0.1442	0.1182	0.1179	0.1210	0.1198	0.7709	0.4717
	$\hat{\beta}_2$	Bias	-0.0152	-0.0149	0.0041	-0.0199	0.0229	0.0133	0.0049	-0.0149	-0.103	-0.7063
		MSE	0.0443	0.0440	0.1301	0.0992	0.1174	0.1130	0.0913	0.0445	0.415	0.2176
500	$\hat{\alpha}$	Bias	0.0001	0.0001	0.0901	-0.0942	-0.002	-0.0001	-0.001	0.003	-0.1491	0.2984
		MSE	0.0002	0.0002	0.1093	0.1892	0.004	0.0154	0.0574	0.0411	0.5395	0.2902
	$\hat{\beta}_1$	Bias	-0.004	-0.0033	0.0074	-0.006	0.0009	0.0006	0.032	-0.054	0.097	0.4496
		MSE	0.0079	0.0077	0.0998	0.0979	0.0036	0.0018	0.0247	0.0183	0.4286	0.3317
	$\hat{\beta}_2$	Bias	-0.0075	-0.006	0.0014	-0.008	0.005	0.007	0.0015	-0.006	-0.1713	-0.4424
		MSE	0.0225	0.0189	0.0938	0.0919	0.0962	0.0685	0.0318	0.0164	0.2774	0.1816

5. Real Data Application

In the past, a nation's overall development levels were determined by its national income because it was believed that the more a nation produced, the more progress it would make both economically and socially. However, we acknowledge that there may be significant differences between societal progress or overall development and GDP growth. Over the past two decades, there has been much discussion about the limitations of using GDP as a gauge of a country's quality of life or social well-being. The fact that a large portion of the population's quality of life has not improved despite a high GDP growth rate has led some people to believe that the GDP measure should be expanded to consider human well-being and life quality.

Unemployment is a critical issue for developing countries because it has a direct and significant impact on a country's economy. It is defined as someone who is willing and able to work but does not have a paid job. Meanwhile, the unemployment rate is the most used indicator for assessing labour market conditions. It is the percentage of people in the labour force who are out of work. Understanding the patterns of unemployment rates is critical these days, and it has piqued the interest of researchers from all fields of study all over the world. For policymakers and researchers, unemployment is important when planning a country's monetary progress.

An advanced modelling approach is required to determine the effect of the unemployment rate efficiently. Several studies have recently relied on traditional testing methods to estimate the effect of the unemployment rate. Furthermore, unemployment is typically non-stationary. As a result, using traditional methods to demonstrate them will yield unpredictable results. To address the issue associated with traditional techniques, a better approach is required to deal with the effect of the unemployment rate (Shi et al., (2022)). The Human Development Index (HDI), a multidimensional indicator of development, has proven to be more reasonable in comparison to the measure of GDP growth, which is one-dimensional in income. This is in line with the general belief that well-being is a multidimensional concept that cannot be measured by market production or GDP alone (Surajit, (2015)), so the value of all goods produced in a nation during a fiscal year is used to define its GDP. It is discovered to be one of the economic growth and production indicators, and to play a crucial strategic role in employment, development, and the balance of payments (Volker, (2005)). In this article, the new procedures were applied to determine the relationships between GDP and HDI. Data were collected from the yearly Jordan's economic report (1990–2021) (Country Economy. Jordan - Human Development Index - HDI 2019 | countryeconomy.com), (Jordan | Data (worldbank.org)) and are presented in Table 7.

Table 7: Yearly Dataset of HDI, GDP and Unemployment Rate of Jordan (1990–2021)

Year	HDI	GDP	Unemployment Rate
1990	0.625	1166.611	16.810
1991	0.636	1155.234	19.513
1992	0.657	1335.288	19.274
1993	0.668	1334.229	19.700
1994	0.679	1414.339	17.171
1995	0.693	1466.045	14.600
1996	0.695	1463.888	13.700
1997	0.699	1494.511	13.686
1998	0.702	1600.398	13.703
1999	0.706	1619.536	13.707
2000	0.711	1651.622	13.700
2001	0.717	1720.361	14.700
2002	0.715	1802.055	15.300
2003	0.720	1876.259	14.500
2004	0.726	2044.964	14.580
2005	0.738	2183.395	14.800
2006	0.741	2513.029	14.000
2007	0.744	2735.379	13.100
2008	0.745	3455.770	12.700
2009	0.743	3559.692	12.900

2010	0.737	3736.645	12.500
2011	0.734	3852.890	12.900
2012	0.735	3910.347	12.200
2013	0.729	4044.427	12.600
2014	0.729	4131.447	11.900
2015	0.730	4164.109	13.080
2016	0.729	4175.357	15.280
2017	0.726	4231.518	18.140
2018	0.728	4308.151	18.270
2019	0.729	4405.487	16.810
2020	0.729	4282.766	19.026
2021	0.730	4405.839	19.252

A descriptive analysis of the data is tabulated in Table 8 with correlations between variables are presented in Table 9. It is noted that there is a strong positive and significant correlation between GDP and HDI ($r = 0.739$, $p < 0.001$) and a strong negative and significant correlation between the unemployment rate and HDI ($r = -0.538$, $p < 0.001$).

Table 8: Descriptive Statistics

Variable	Min	Max	Mean	STDEV
Unemployment Rate	11.9	19.7	15.1	2.5
GDP	1155.2	4405.8	2726.3	1242.6
HDI	.63	.75	.71	.03

Table 9: Correlation Matrix between the dependent variable (HDI) and independent variables (GDP and Unemployment rate).

Variable	HDI
GDP	$.739^{**}$ (0.000)
Unemployment Rate	$-.538^{**}$ (0.001)

Note: ** significant at 0.01 level. Values in () represent the p-value.

The trend of the variables within the study period is given in Figures 3, 4 and 5.

Pak.j.stat.oper.res. Vol.20 No. 3 2024 pp 471-488 DOI: http://dx.doi.org/10.18187/pjsor.v20i3.4483

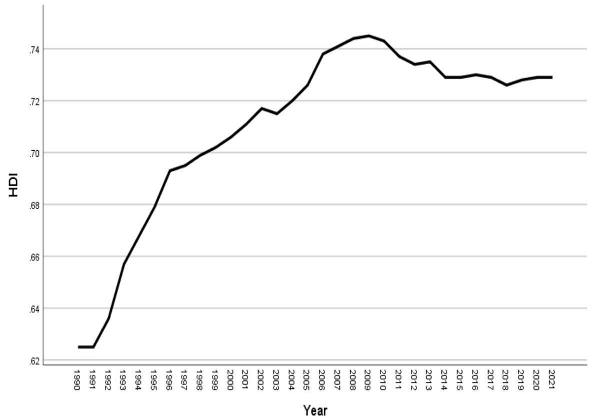

Figure 3: The trend of the HDI within 1990-2021.

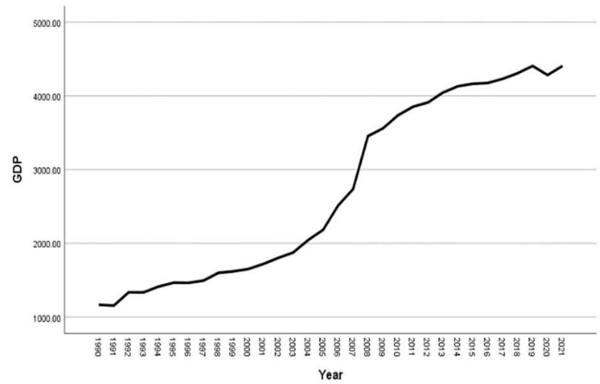

Figure 4: The trend of the national GDP within 1990-2021.

Figure 5: The trend of the unemployment rate within 1990-2021

Moreover, the scatter plots in Figures 6 and 7 suggest that there is almost a linear relationship between the variables, Also, based on the linearity test, the result indicates there is a weak linear relationship between the variables (F = 0.316, P.=0.98), as shown in Figure 8. Moreover, the scatter plots in Figure. 9. indicate that there is a heteroscedasticity problem in fitting the model.

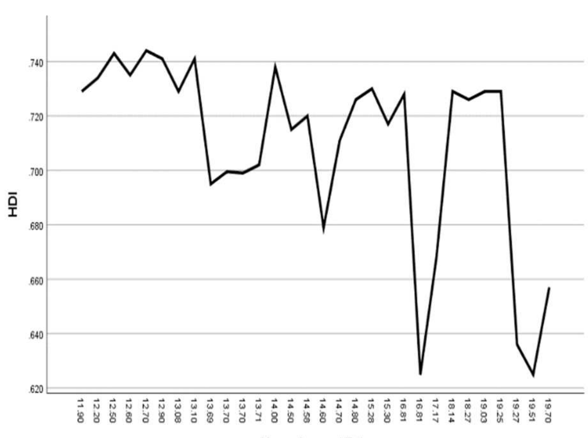

Figure 6: The line plot of HDI and Unemployment rate.

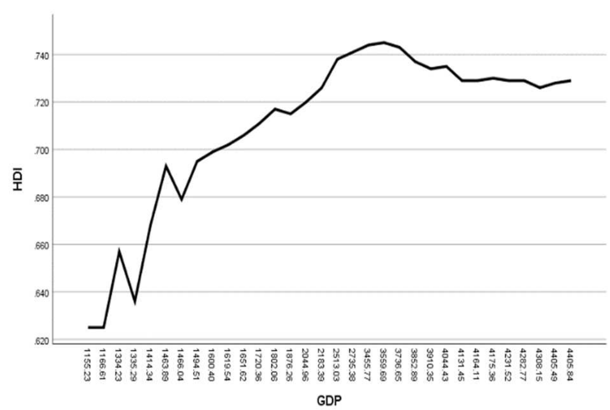

Figure 7: The line plot of HDI and GDP

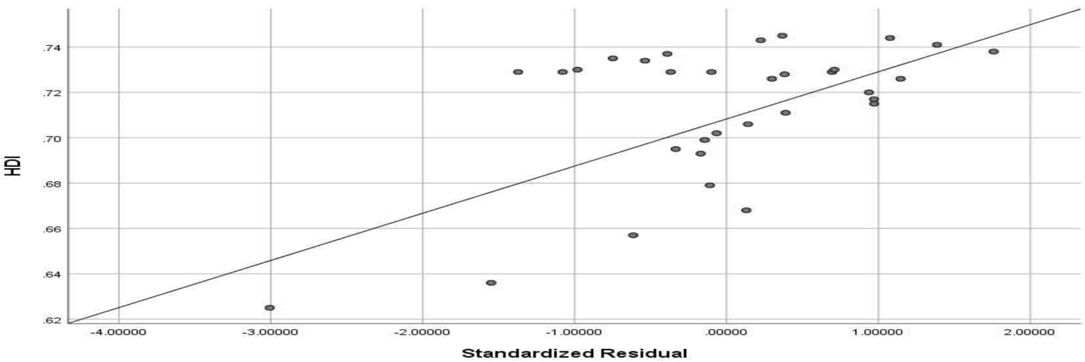

Figure 8: The residual plot

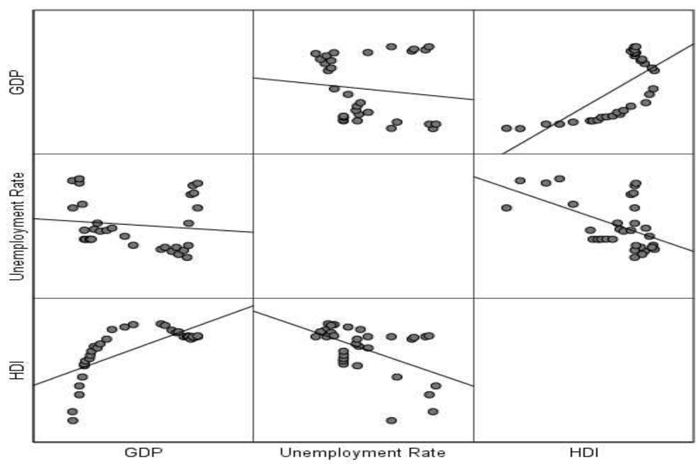

Figure 9: The scatter plot of HDI, GDP and unemployment rate

These analyses suggest that the GDP, unemployment rate, and HDI can be modelled as linear relationships; however, it is believed that all variables are subject to error because their value is affected by several other factors. As a result, it is suggested to consider MEM for studying the relationship between HDI, unemployment rate, and GDP. The model under consideration can therefore be reformulated as follows:

$$\text{HDI} = \alpha + \beta_1 \times (\text{GDP} - \delta_1) + \beta_2 \times (\text{Unemployment Rate} - \delta_2) + \epsilon. \quad (18)$$

Table 10 displays the outcomes of the estimation methods considered in this article: The Repetitive Weighted, Iterative weighted, MLE and MOM. The results indicate that, based on mean square residual (MSR), the proposed methods with r =3 and r =4 produced more accurate estimators than the other estimation methods in weight case 1, Additionally, the proposed method with r=4 provided more accurate estimators than the other methods in weight case 2. Furthermore, the results show that the proposed estimators using weighted technique 1, outperformed those using technique 2. Overall, the proposed procedures demonstrated greater efficiency than MLE and MOM in fitting the model, as illustrated in Figures 10 and 11, which display the residuals for each estimation method.

Table 10: Parameter Estimation of HDI vs GDP and Unemployment rate

Weight case	Method	Criterion	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	MSR
1	Repetitive weighted	r = 3	5e-06	0.0009	0.6658	0.00095
		r = 4	5e-06	0.0009	0.6658	0.00095
	Iterative weighted	r = 3	0.0002	0.0459	-0.4868	0.0527

		r = 4	0.0003	0.0461	-0.6882	0.0949
2	Repetitive weighted	r = 3	0.0002	0.0199	-0.2246	0.0589
		r = 4	0.0001	0.0202	0.0478	0.0160
	Iterative weighted	r = 3	0.0002	0.0244	-0.2951	0.0801
		r = 4	0.0001	0.0433	-0.3008	0.0334
Classical	MLE		1.6e-05	0.0026	0.651	0.1926
	MOM		1.15e-05	0.00624	0.3481	0.1083

Note: MSR = $\frac{\sum_{i=1}^{n}\varepsilon^2}{n} = \frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}$, where, $\hat{y} = \hat{\alpha} - \bar{x}_1\hat{\beta}_1 - \bar{x}_2\hat{\beta}_2$, y is a HDI observations.

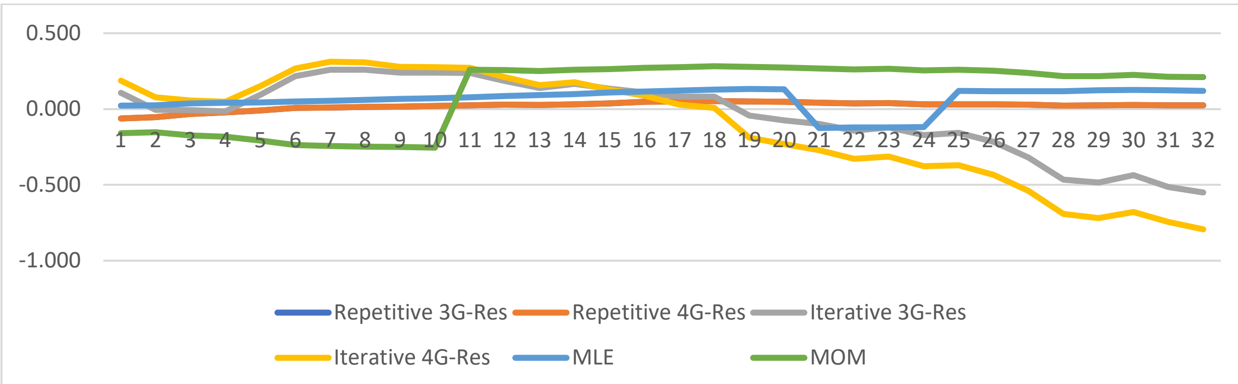

Figure 10: Residual of each estimation method for Case 1

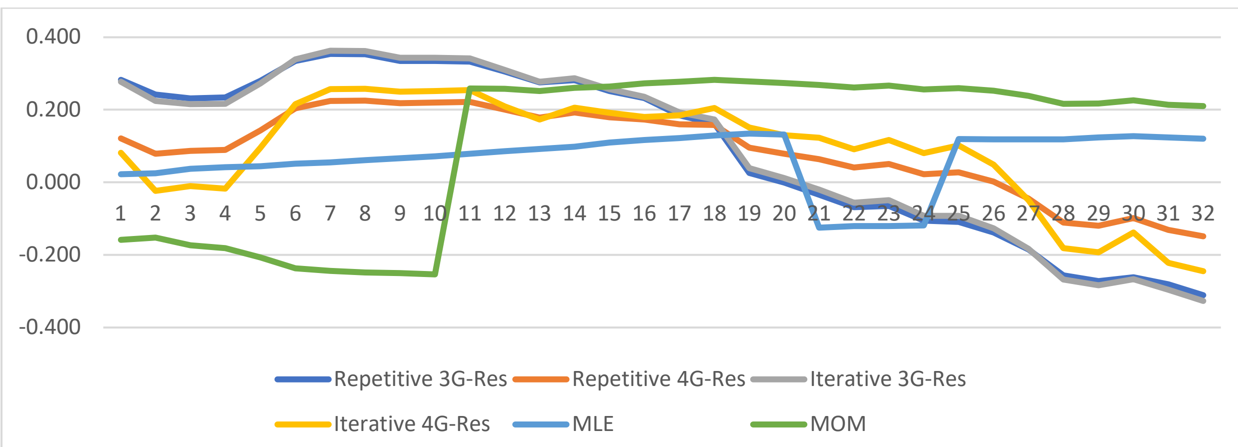

Figure 11: Residual of each estimation method for Case 2

6. Concluding Remarks

This study proposed two new nonparametric estimation procedures for fitting multiple structural MEMs: The Repetitive Weighted Grouping procedure and the Iterative Weighted Grouping procedure. Monte Carlo simulations illustrate the superiority of the proposed estimation procedures over the classical methods (MLE and MOM) across various sample sizes. Additionally, the results indicate that both proposed procedures perform better in weight case 2 compared to weight case 1, highlighting their greater efficiency in fitting multiple structural MEMs. A key contribution of this work is the introduction of these new techniques, which were not covered in earlier studies. When comparing these techniques to those from prior research (by Ahmad and Ahmad, (2019), Salem, (2018), Schennach and Hu, (2013), Cheng and Van Ness, (1999), Adusumilli and Otsu, (2018), and Fuller, (1987)) that used MLE and MOM approaches, the proposed methods showed improved performance. Furthermore, real data analysis was conducted to explore the effects of GDP and the unemployment rate on the HDI. The findings revealed a strong positive relationship between GDP and HDI, while a strong negative relationship was observed between the unemployment rate and HDI.

References

1. Adusumilli, K. and Otsu, T. (2018). Nonparametric Instrumental Regression with Errors in Variables. Econometric Theory, 34(6), 1-25. DOI:10.1017/S0266466617000469.
2. Ahmad, A., & Ahmad, S. P. (2019). Weighted Analogue of Inverse Gamma Distribution: Statistical Properties, Estimation and Simulation Study. Pakistan Journal of Statistics and Operation Research, 15(1), 25-37. https://doi.org/10.18187/pjsor.v15i1.2238.
3. Al-Dibi'i, R. and Al-Nasser, A. D. (2019). Fitting Structural Measurement Error Models Using Repetitive Wald-Type Procedure. The 6th International Arab Conference on Mathematics and Computations, 24-26, April 2019, Al Zarqa University: Jordan, 6(1), 182-187.
4. Al-Nasser, A. D. (2004). Estimation of Multiple Linear Functional Relationships. Journal of Modern Applied Statistical Methods, 3(1),181-186. doi: 10.22237/jmasm/1083370920.
5. Al-Nasser, A. D. (2005). Entropy Type Estimator to Simple Linear Measurement Error Models. Austrian Journal of Statistics, 34 (3), 283-294. DOI:10.17713/ajs.v34i3.418.
6. Al-Nasser, A. D. (2011). An Information-Theoretic Alternative to Maximum Likelihood Estimation Method in Ultrastructural Measurement Error Model. Hacettepe Journal of Mathematics and Statistics, 40(3), 469 – 481.
7. Al-Nasser, A. D. (2012). On using the Maximum Entropy Median for Fitting the Un-replicated Functional Model Between the Unemployment Rate and the Human Development Index in the Arab Stats. Journal of Applied Sciences, 326-335. DOI: 10.3923/jas.2012.326.335.
8. Al-Nasser, A. D., Al-Sliti, A. and Edous, M. (2016). New iterative AM estimation procedure for fitting the simple linear measurement error models. Electronic Journal of Applied Statistical Analysis, 9(3), 491-501. DOI:10.1285/i20705948v9n3p491.
9. Bartlett, M. S. (1949). Fitting a straight line when both variables are subject to error. Biometrics, 5, 207-2012.
10. Cabral, C. R. B., Souza, N. L. and Leão, J. (2020). Bayesian Measurement Error Models Using Finite Mixtures of Scale Mixtures of Skew-Normal Distributions. https://arxiv.org/abs/2007.13037v1. https://doi.org/10.48550/arXiv.2007.13037.
11. Cao, C., Chen, M. and Wang, Y. (2018). (b) Heteroscedastic replicated measurement error models under asymmetric heavy-tailed distributions. Computational Statistics, 33(1), 319 – 338. DOI: 10.1007/s00180-017-0720-8.
12. Cao, C., Chen, M., Ren, Y. and Xu, Y. (2018). (a) Robust replicated heteroscedastic measurement error model using heavy-tailed distribution. Communications in Statistics - Simulation and Computation, 47(6), 1771-1784. https://doi.org/10.1080/03610918.2017.1324979.
13. Carroll, J. R., Ruppert, D. A., Stefanski L. & Crainiceanu, M. C. (2006). Measurement Error in Nonlinear Models, NY: Chapman and Hall/CRC. https://doi.org/10.1201/9781420010138.
14. Carroll, R. J., Delaigle, A. & Hall, P. (2007). Non-parametric regression estimation from data contaminated by a mixture of Berkson and classical errors. Journal of the Royal Statistical Society: Series B, 69(5), 859–878. doi: 10.1111/j.1467-9868.2007.00614.x
15. Cheng, C. L. and Van Ness, J. W. (1999). Statistical Regression with Measurement Error. London: Arnold.
16. Country Economy. Jordan - Human Development Index - HDI 2019 | countryeconomy.com.
17. Cragg, J.G. (1997). Using higher moments to estimate the simple errors-in-variables model. The RAND Journal of Economics, 28(0), S71-S91. https://doi.org/10.2307/3087456.
18. Drion, E.F. (1951). Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. Indagationes Mathematicae, (54), 256-260. https://doi.org/10.1016/S1385-7258(51)50036-7.
19. Dunn, G. (2000). Statistical Evaluation of Measurement Errors (2nd Edition). Arnold, London.
20. Fuller, W. A. (1987). Measurement Error Models, New York: Wiley. DOI:10.1002/9780470316665
21. Geary, R. C. (1942). Inherent relations between random variables. In Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences, 47, 63–76. https://www.jstor.org/stable/20488436
22. Gillard, J. (2010). An Overview of Linear Structural Models in Errors in Variables Regression. Revstat-Statistical Journal, 8(1), 57-80. DOI:10.57805/revstat.v8i1.90.
23. Gillard, J.W. and Iles,T.C. (2006). Variance covariance matrices for linear regression with errors in both variables. Cardiff School of Mathematics Technical Report.

24. Green, W.H. (2011). Econometric Analysis, 7th edition, Prentice-Hall.
25. Gupta, Y. P., and Amanullah. (1970). A Note on the Moments of the Wald's Estimator. Statistica Neerlandica 24(3), 109-123. https://doi.org/10.1111/j.1467-9574.1970.tb00113.x.
26. Kendall, M. and Stuart, A. (1979). The advanced theory of statistics. Inference and relationship. London: Griffin, 2, 4th ed. DOI: 10.2307/2985294
27. Lindley, V. D. (1947). Regression Lines and the Linear Functional Relationship. Journal of the Royal Statistical Society, 9(2), 218-244.
28. Madansky, A. (1959). The Fitting of Straight Lines when Both Variables are Subject to Error. Journal of the American Statistical Association, 54(285), 173-205. https://doi.org/10.1080/01621459.1959.10501505.
29. Nair, R. K. and Shrivastava, P. M. (1942). On a Simple Method of Curve Fitting. Sankhyā: The Indian Journal of Statistics, 6 (2), 121-132.
30. Pal, M. (1980). Consistent moment estimators of regression coefficients in the presence of errors in variables. Journal of Econometrics, 14(3), 349–364. https://doi.org/10.1016/0304-4076(80)90032-9
31. Raj. B. and Koerts, J. (1992). Henri Theil's Contributions to Economics and Econometrics, (23), 1570-5811.
32. Salem, H. M. (2018). Parameter Estimations and Optimal Design of Simple Step-Stress Model for Gamma Dual Weibull Distribution. Pakistan Journal of Statistics and Operation Research, 14(1), 109-120. https://doi.org/10.18187/pjsor.v14i1.1671.
33. Schennach, S. M. and Hu, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. Journal of the American Statistical Association, 108(501), 177–186. http://www.jstor.org/stable/23427520.
34. Shi, L., Khan, YA. and Tian, M-W. (2022). COVID19 pandemic and unemployment rate prediction for developing countries of Asia: A hybrid approach. PLOS ONE, 17(12), e0275422. https:// doi.org/10.1371/journal.pone.0275422.
35. Spicker, D. M. Wallace, P. and Yi, G. Y. (2021). Nonparametric Simulation Extrapolation For Measurement Error Models. arXiv:2111.02863v1. https://doi.org/10.48550/arXiv.2111.02863.
36. Surajit, D. (2015). Gap between GDP and HDI: Are the Rich Country Experiences Different from the Poor?IARIW-OECD Special Conference: "W(h)ither the SNA? Paris, France.
37. The World Bank. Jordan | Data (worldbank.org).
38. Van Montfort, K. (1988). Estimating in Structural Models with Non-Normal Distributed Variables: Some Alternative Approaches. M & T Series 12. DSWO Press, Leiden.
39. Volker, T. (2005). Tanzania's Growth Process and Success in Reducing Poverty. IMF Working Paper.
40. Wald, A. (1940). The Fitting of Straight Lines if Both Variables are Subject to Error. Ann. Math. Statist, 11 (3), 284-300. DOI: 10.1214/aoms/1177731868.
41. Wiedermann, W., Merkle, E. C. and von Eye, A. (2018). Direction of dependence in measurement error models. British Journal of Mathematical and Statistical Psychology, 71, 117– 145. DOI: 10.1111/bmsp.12111.
42. Xu, Y., Kim, J. K. and Li, Y. (2017). Semiparametric estimation for measurement error models with validation data. Canadian Journal of Statistics, 45(2), 185–201. https://doi.org/10.1002/cjs.11314.
43. Yi, G. Y. (2017). Statistical Analysis with Measurement Error or Misclassification. Springer New York.