



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

On Using the Maximum Entropy Median for Fitting the Unreplicated Functional Model Between the Unemployment Rate and the Human Development Index in the Arab States

Amjad D. Al-Nasser

Department of Economics and Statistics, College of Business Administration,
 University of Dubai, Dubai, UAE

Abstract: In this study modified estimation methods based on the maximum entropy median for fitting unreplicated functional model to data were proposed. Two types of modification were considered, modified Wald-type estimators and modified Theil-type estimators. Also, a coefficient of determination to estimate the proportion of variation when both variables are measured with error is proposed. A Monte Carlo experiment were used to study the performance of the proposed estimators, the simulation results showed that the proposed estimators are more efficient than the classical estimators when the data contains outliers. A real data analysis for fitting the relationships between the unemployment rate and the human development index in the Arab's states also included. The data analysis showed that there is a negative relationship between both variables and suggested that the Arab's states should change their economical, educational and social policies to improve the life quality in these states.

Key words: Maximum entropy median, perpendicular error, outliers, unemployment rate, human development index

INTRODUCTION

The classical method used in fitting simple linear relationship between two variables; y and x , is the Ordinary Least Squares (OLS) method by minimizing the sum of squares of the vertical distances “regression y on x ” or the horizontal distances “regression x on y ”; of the observations from the straight line; assuming that only one variable is measured with error. In fact, most of the analytical methods are subject to measurement errors and ignoring such error then not only the slope attenuated but also the data are noisier; with an increased error about the line (Carroll *et al.*, 2004). Therefore, we should consider the migrations in x by distinguish between the measured value (x) and the true value:

$$\xi = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n X_i}{n}$$

by considering the simple linear Measurement Error Model (MEM) instead of the simple linear regression. Then; for a sample of size n ; the simple MEM between two unobserved latent variables ξ and η are assumed to have the following relationships:

$$\eta_i = \alpha + \beta \xi_i, I = 1, 2, \dots, n \quad (1)$$

To measure such latent variables with additive error one uses manifest variable Y and X , then the observed data are:

$$y_i = \eta_i + \varepsilon_i \text{ and } x_i = \xi_i + \delta_i \quad (2)$$

where, δ_i and ε_i are assumed to be mutually independent with mean $(0,0)$ and unknown variances $(\sigma_\delta^2, \sigma_\varepsilon^2)$. There are $(n+4)$ parameters to be estimated in (1) and (2) which are α , β , σ^2 , τ^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. To overcome the inconsistencies of the estimators, one piece of information is required on the MEM (Kendall and Stuart, 1979; Fuller, 1987; Cheng and Van Ness, 1999; Al-Nasser *et al.*, 2005; Carroll *et al.*, 2009; Fah *et al.*, 2007; O'Driscoll and Ramirez, 2011). In model (1) and (2) the vertical and the horizontal distances should be minimized at once in order to estimate the unknown parameters and this can be done by minimizing the orthogonal distances (Fig. 1) or the so called squared perpendicular error:

$$\sum_{i=1}^n \left(\frac{y_i - \alpha - \beta x_i}{\sqrt{1 + \beta^2}} \right)^2$$

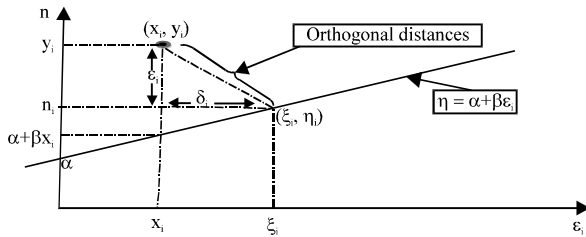


Fig. 1: Orthogonal distances from the data to the line

Under normality assumptions, the distances in Fig. 1 are given as follows:

$$\delta_i = \frac{\beta}{\lambda + \beta^2} (y_i - \alpha - \beta x_i)$$

$$\epsilon_i = \frac{\lambda}{\lambda + \beta^2} (y_i - \alpha - \beta x_i)$$

where,

$$\lambda = \frac{\sigma_\epsilon^2}{\sigma_\delta^2}$$

The errors then could be estimated by the classical OLS or maximum likelihood estimation methods given that λ is known and in case it is unknown (Jonathan and Terence, 2009) proposed an estimator of it based on the fourth moments:

$$\hat{\lambda} = \frac{S_{yy} - S_{xy} \sqrt{\frac{S_{xyyyy} - 3S_{xy}S_{yy}}{S_{xxxx} - 3S_{xy}S_{xx}}}}{S_{xx} - S_{xy} \sqrt{\frac{S_{xyyyy} - 3S_{xy}S_{yy}}{S_{xxxx} - 3S_{xy}S_{xx}}}} \quad (3)$$

where,

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2, S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2, S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{yyyy} = \sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2, S_{xxxx} = \sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})$$

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

As a matter of fact, the difficulties appeared in the OLS and the maximum likelihood estimation methods are

good reasons to use alternative methods, the most efficient and easy methods in the literature are Wald-type (Wald, 1940; Al-Radaideh *et al.*, 2010; Huwang and Yang, 2000; Hussin *et al.*, 2009; Al-Nasser *et al.*, 2005) and Theil-type estimate based on repeated median methods (Theil, 1950; Sen, 1968; Siegel, 1982).

In this study, a modified Wald-type and Theil-type estimators proposed based on the maximum entropy median which are preferable to get new robust estimators. Moreover, a coefficient of determination for the unrepliated functional model is proposed.

ESTIMATION METHODS OF UNREPLICATED FUNCTIONAL MODEL (URFM)

Various estimation methods have been suggested to estimate the unknown parameters of the URFM (Hussin *et al.*, 2009; Al-Nasser *et al.*, 2005). Two different types of estimation methods will be considered in this article, mean-based methods, i.e., Maximum Likelihood Estimation (MLE) and Grouping Estimation methods and median based methods, i.e., Theil-type estimation methods.

Maximum likelihood estimators: The classical MLE of model (1) and (2) can be obtained by minimizing the Mahalanobis distance:

$$\frac{\sum_{i=1}^n (x_i - \xi_i)^2}{2\sigma_\delta^2} + \frac{\sum_{i=1}^n (y_i - \alpha - \beta \xi_i)^2}{2\sigma_\epsilon^2}$$

Under the normality assumption; the log likelihood function of the URFM:

$$\log L(\alpha, \beta, \sigma_\delta^2, \sigma_\epsilon^2, \xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n) = -n \log(2\pi) - \frac{n}{2} (\log(\sigma_\delta^2) + \log(\sigma_\epsilon^2)) - \frac{\sum_{i=1}^n (x_i - \xi_i)^2}{2\sigma_\delta^2} + \frac{\sum_{i=1}^n (y_i - \alpha - \beta \xi_i)^2}{2\sigma_\epsilon^2}$$

Which can be rearrange in the following form:

$$L = \text{constan t} - n \ln(\sigma^2) - \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (x_i - \xi_i)^2 + \lambda^{-1} \sum_{i=1}^n (y_i - \alpha - \beta \xi_i)^2 \right\}$$

Solari (1969) showed that there is no MLE solution for the slope if λ is unknown. Therefore, the piece of

information about λ (or both error variances are known) is essential to obtain a consistent estimator (Kendall and Stuart, 1979). Then, based on the given information the MLE solution will be:

$$\hat{\beta} = \frac{(S_{yy} - \lambda S_{xx}) + ((S_{yy} - \lambda S_{xx})^2 + 4\lambda S_{xy})^{1/2}}{2S_{xy}} \quad (4)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

and:

$$\hat{\xi}_i = \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{\lambda + \hat{\beta}^2}$$

Then the corrected perpendicular distance (Cheng and Van Ness, 1999) can be obtained by:

$$\hat{\sigma}_e^2 = \frac{\lambda}{2n(\lambda + \hat{\beta}^2)} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}\hat{\xi}_i)^2$$

It should be noted that; when $\lambda = 1$, the URFM is known as Deming regression. Also, the URFM estimate of the slope in Eq. 4 can be expressed as a combination of the regressions of y on x and x on y , which can be estimated by:

$$\hat{\beta} = d + \text{sgn}(S_{xy})(d^2 + 1)^{1/2}$$

where,

$$d = 1/2(\hat{\beta}_{xy} - \lambda/\hat{\beta}_{yx}) \text{ and } \hat{\beta}_{xy} = S_{xy}/S_{xx}, \hat{\beta}_{yx} = S_{xy}/S_{yy}$$

Noting that; $\hat{\beta}_{yx}$ and $\hat{\beta}_{xy}$ are the regression estimate of y on x and x on y , respectively.

Wald estimators: One of the most popular alternative methods for MLE in the literature is the grouping method and well-known as Wald-type estimators. The method consists of ordering the observed pairs (x_i, y_i) 's by the magnitude of the x_i 's, then divide the observations into some groups. Wald (1940) suggested two groups while Nair and Shrivastava (1942) and Bartlett (1949), proposed a more efficient estimators by forming the observations into three groups; such that the estimate will be based on the first and the last group while middle group will be omitted. In general, the grouping method has the following procedure (Al-Nasser *et al.*, 2005):

- Order the pairs (x_i, y_i) s by the magnitude of x_i ;
- Select proportions P_1 and P_2 such that $P_1 + P_2 < 1$, place the first nP_1 pairs in group one (G_1) and the last nP_2

pairs in another group (G_3), discarding G_2 the middle group of observations that is to say:

$$\text{Place}(x_i, y_i) \text{ in } \begin{cases} G_1 & \text{if } x_i \leq x_{p_1} \\ G_2 & \text{if } x_{p_1} < x_i \leq x_{1-p_2} \\ G_3 & \text{if } x_i > x_{1-p_2} \end{cases}$$

where, x_{p_i} is the p_i percentile

- Then the slope and the intercept of URFM can be estimated, respectively by:

$$\hat{\beta} = \frac{P_2^{-1} \sum_{G_2} Y_{rp2(t)} - P_1^{-1} \sum_{G_1} Y_{rp1(t)}}{P_2^{-1} \sum_{G_2} Y_{rp2(t)} - P_1^{-1} \sum_{G_1} Y_{rp1(t)}}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

where, $\{X_{rp(i)}, I = 1, 2, \dots, n\}$ are the order statistics of X and $\{Y_{rp(i)}, I = 1, 2, \dots, n\}$ are the associated Y . Noting that when:

$$P_1 = P_2 = \frac{1}{2}$$

then the grouping method named by two groups and when:

$$P_1 = P_2 = \frac{1}{3}$$

then the method is called three groups

The advantage of these methods is the simplicity in computation and avoiding of the normality assumption. However, some weak precision of using this method arises if there is a set of outliers, because it is mean-based estimators. To avoid this problem it is preferable to use the median as center of gravity of each subgroup; then the Wald-type median based (two groups) estimators will be:

$$\hat{\beta} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} \text{ and } \hat{\alpha} = \text{med}(y_t - \hat{\beta}x_t) \quad (5)$$

where, $\bar{y}_j = \text{med}(y_i)$ and $\bar{x}_j = \text{med}(x_i)$ are the medians of subgroup j , $I = 1, 2, \dots, n$, $j = 1, 2$.

The similar modification can be made with three group's method.

Robust estimators: Theil (1950) proposed a robust method for making inference about the slope of the unreplicated functional model given in Eq. 1 and 2. The method is based on finding the median of all possible pairs of observations, assuming that all x_i 's are distinct:

$$B_{ij} = \frac{y_j - y_i}{x_j - x_i}, i, j = 1, 2, \dots, ni < j \quad (6)$$

Sen (1968) extended the proposed estimator in Eq. 6 to the case in which two samples have the same x-coordinates by considering only the slope values with distinct x-coordinates. Then the URFM parameters can be estimated by:

$$\hat{\beta} = \text{med}(B_{ij}) \quad (7)$$

$$\hat{\alpha} = \text{med}(y_i - \hat{\beta}x_i) \quad (8)$$

However, Maritz (1979), extending Theil's ideas by proposing a point estimate for the intercept term, by taking the median of all pairwise intercept estimates:

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

Hence:

$$\hat{\alpha} = \text{med}(A_{ij}) \quad (9)$$

Moreover, Siegel (1982) introduced another estimate based on repeated median method. The slope of URFM can be estimated by two stages after evaluating Theil statistics given in Eq. 6:

- **First stage:** $\hat{\beta}_i = \text{med}(B_{ij})$; $i, j = 1, 2, \dots, n-1$ (10)

- **Second stage:** $\hat{\beta} = \text{med}(\hat{\beta}_i)$ (11)

Then, the intercept can be estimated by using the same formula given in Eq. 8 or 9.

PROPOSED ESTIMATORS

Different types of estimators are proposed for the URFM. The first one is a modified Wald-type and Theil-Type parameter estimators based on the maximum entropy median and the second one is a coefficient of determination to estimate the proportion of variation in y that explained by x when both variables are measured with error.

Modified parameter estimators for URFM: As an alternative to the sample median; Theil and O'Brien (1980) proposed the ME median which is defined as:

$$\text{ME.median} = \begin{cases} \frac{1}{2}(x_{(n/2)} + x_{((n/2)+1)}) & , n \text{ even} \\ \frac{1}{4}(x_{((n-1)/2}) + \frac{1}{2}(x_{(n+1)/2}) + \frac{1}{4}(x_{((n+3)/2})) & , n \text{ odd} \end{cases}$$

it is obvious that for even sample size (n) the ME median is equal to the sample median. In case of add sample size; Bani-Mustafa and Al-Nasser (2008), showed that the ME-median is more robust than the sample median with small sample size and performs almost equally well as the sample median with large samples. These results could be considered as good evidence in order to use the maximum entropy median instead the sample median to estimate the URFM parameters. Based on this, there are two modified estimators, the first is to use the ME median in the Wald-type estimators, hence the first modification will be on the estimators given in Eq. 5 to be:

$$\hat{\beta} = \frac{\text{ME.median}(y_2) - \text{ME.median}(y_1)}{\text{ME.median}(x_2) - \text{ME.median}(x_1)}$$

and:

$$\hat{\alpha} = \text{ME.median}(y_i - \hat{\beta}x_i)$$

The second modification is about the Theil-Type estimators and it is divided into two groups:

- **Group 1:** Modification of Eq. 7 and 8:

$$\hat{\beta} = \text{ME.median}(B_{ij})$$

$$\hat{\alpha} = \text{ME.median}(y_i - \hat{\beta}x_i)$$

- **Group 2:** Modification of Siegel estimators given in Eq. 10 and 11:

- **First stage:** $\hat{\beta}_i = \text{ME.median}(B_{ij})$ $i = 1, 2, \dots, n-1$
- **Second stage:** $\hat{\beta} = \text{ME.median}(\hat{\beta}_i)$

Then, the intercept can be estimated by using the modified formula of Eq. 8 or 9, where, B_{ij} is given in Eq. 6.

Proposed coefficient of determination for the URFM: Coefficient of determination is a measure used

in statistical model analysis to assess how well a model explains and predicts future outcomes on the basis of other related information. It is the proportion of variability in a data set that is accounted for by the statistical model in order to measure the accuracy of the model. In the context of simple linear regression this coefficient is estimated as:

$$\frac{S_{xy}^2}{S_{xx} S_{yy}}$$

For the URFM there is no such measure in the literature, however, this measure can be derived based on the second sample moments which converge in probability to their expectations:

$$\begin{aligned} S_{xx} & \xrightarrow{p} S\xi\xi + \sigma_\xi^2 \\ S_{yy} & \xrightarrow{p} \beta^2 S\xi\xi + \sigma_\varepsilon^2 \\ S_{xy} & \xrightarrow{p} \beta S\xi\xi \end{aligned}$$

Therefore, when both variables are measured with errors; the population coefficient of determination is given by:

$$R_{URFM}^2 = \frac{(\beta S\xi\xi)^2}{(S\xi\xi + \sigma_\xi^2)(\beta^2 S\xi\xi + \sigma_\varepsilon^2)}$$

Alternatively, this coefficient can be estimated by the unexplained variance based on the relationships between the Sum of Squared Errors (SSE) and the regression sum of squares (SSR). The SSE in the URFM can be estimated by:

$$\frac{n}{n-1} (S_{yy} - \hat{\beta} S_{xy})$$

however, this estimator is not very reliable (Cheng and Van Ness, 1999). Instead, we can use the perpendicular error:

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{\lambda}{2n(\lambda + \hat{\beta}^2)} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \hat{\xi}_i)^2 \\ &= \frac{\hat{\lambda}}{2(\hat{\lambda} + \hat{\beta}^2)} (S_{yy} - 2\hat{\beta} S_{xy} + \hat{\beta}^2 S_{xx}) \end{aligned} \tag{12}$$

Moreover; and based on different estimation methods; the estimator given in Eq. 12 is inconsistent and should be adjusted based on the degrees of freedom i.e., n-2. Then the consistent estimator for SSE will be:

$$SSE = \frac{n\hat{\lambda}}{(n-2)(\hat{\lambda} + \hat{\beta}^2)} (S_{yy} - 2\hat{\beta} S_{xy} + \hat{\beta}^2 S_{xx})$$

Consequently, the URFM regression sum of squares (SSR) can be derived from the differences between the total sum of squares (S_{yy}) and SSE:

$$SSR = S_{yy} - SSE$$

Hence, the estimated coefficient of determination i.e., proportion of variation in y that explained by x when both variables are measured with error; can be estimated by:

$$\hat{R}^2 = \frac{SSR}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$$

Performance of the proposed parameter estimators:

Monte Carlo experiments were conducted to evaluate the performance of the new proposed estimators based on maximum entropy median. The experiments were carried out by generating 10000 samples from the simple SLFR model:

$$\begin{aligned} y_i &= 1 + x_i + \varepsilon_i \\ x_i &= \xi_i + \delta_i, \quad I = 1, 2, \dots, n \end{aligned}$$

assuming that; $\varepsilon \sim N(0, 1)$, $\delta \sim N(0, 1)$ and:

$$\xi_i = -\left[\frac{n}{2}\right], \dots, \left[\frac{n}{2}\right],$$

where, [*] is the largest number less than or equal to *.

In order to study the robustness of the proposed estimator; the sample sizes were selected to be 15, 35 and 95 to ensure odd subgroup sizes for the Wald-type estimators. Moreover, we considered the contaminated case or the so called slippage in variance model, for each sample, exactly one of the observed data points x_i 's (y_i 's) is drawn from $N(0, 4^2)$ and n-1 observations were drawn from the actual distribution i.e., $N(0, 1^2)$.

Moreover, the outlier observed values selected into two cases, the first one to be fixed; by considering the position of the outlier to be the last data point and the other selection of the outlier position to be random in the interval (1,n). For the MLE based estimators, we assumed that $\lambda = 1$ is known (Deming regression). Then the performance of all estimates evaluated measuring the efficiency of the estimators:

$$Eff = \frac{MSE(\hat{p}_{MLE})}{MSE(\hat{p}_*)}$$

Table 1: Efficiency of the SLFR model with contamination in X

| Sample size | Estimation method | No outliers | | Outliers | | |
|-------------|-------------------|-------------|---------|----------|---------|------|
| | | α | β | α | β | |
| 15 | 2G | Mean | 1.00 | 0.80 | 0.96 | 1.22 |
| | | Median | 1.83 | 0.71 | 0.89 | 1.64 |
| | | ME-Median | 1.41 | 0.83 | 0.97 | 1.82 |
| | 3G | Mean | 0.98 | 0.89 | 0.99 | 1.03 |
| | | Median | 0.79 | 0.18 | 1.16 | 1.09 |
| | | ME-Median | 0.58 | 0.47 | 1.25 | 1.47 |
| | Theil | Median | 0.54 | 0.88 | 1.15 | 3.20 |
| | | ME-Median | 0.55 | 0.89 | 1.52 | 3.23 |
| | Siegel | Median | 0.64 | 0.56 | 1.42 | 2.86 |
| ME-Median | | 0.67 | 0.60 | 1.44 | 2.97 | |
| 35 | 2G | Mean | 1.00 | 0.76 | 1.00 | 0.84 |
| | | Median | 0.60 | 0.82 | 1.04 | 1.20 |
| | | ME-Median | 0.63 | 0.81 | 1.05 | 1.23 |
| | 3G | Mean | 1.00 | 0.95 | 0.99 | 0.97 |
| | | Median | 0.58 | 0.70 | 1.55 | 1.30 |
| | | ME-Median | 0.64 | 0.81 | 1.63 | 1.42 |
| | Theil | Median | 0.65 | 0.93 | 4.93 | 4.73 |
| | | ME-Median | 0.67 | 0.93 | 4.93 | 4.76 |
| | Siegel | Median | 0.68 | 0.51 | 5.30 | 4.31 |
| ME-Median | | 0.68 | 0.56 | 5.33 | 4.41 | |
| 95 | 2G | Mean | 0.99 | 1.06 | 0.96 | 2.07 |
| | | Median | 0.95 | 0.41 | 1.15 | 1.49 |
| | | ME-Median | 0.96 | 0.54 | 1.53 | 1.74 |
| | 3G | Mean | 0.99 | 1.18 | 0.95 | 1.01 |
| | | Median | 1.46 | 0.69 | 1.43 | 2.98 |
| | | ME-Median | 1.49 | 0.84 | 1.44 | 3.36 |
| | Theil | Median | 2.91 | 4.31 | 1.60 | 3.80 |
| | | ME-Median | 2.95 | 4.33 | 1.62 | 3.83 |
| | Siegel | Median | 3.08 | 4.60 | 1.56 | 3.52 |
| ME-Median | | 3.11 | 4.64 | 1.57 | 3.56 | |

*2G: Two groups and 3G: Three groups

where, ρ could be the intercept or the slope and $\hat{\rho}$, could be any of the estimators considered in this article. The simulation results reported in Table 1 and 2 showed that the ME-Median based estimators are more efficient than MLE, Wald-type or Theil-Type estimators whenever the data contains an outlier. However, for the inliers data generation the MLE based estimators are more robust.

EFFECTS OF UNEMPLOYMENT RATE ON HUMAN DEVELOPMENT INDEX

The Unemployment Rate (UR) is an indicator of employable people in a country's workforce who are over the age of 16 and who have either lost their jobs or have unsuccessfully sought jobs in the last month and are still actively seeking work (Kayani and Ghani, 2003; Tan and Santhapparaj, 2007; Fazoranti, 2009).

Statistically, UR is computed as a ratio of the unemployed workers to the total labor force. Lately, unemployment or joblessness is considered one of the major problems in the Arab states and only three countries (UAE, Qatar and Kuwait) out of the 22 formal members state of the Arab league (Appendix 1), have UR

Table 2: Efficiency of the SLFR model with contamination in Y

| Sample size | Estimation method | No outliers | | Outliers | | |
|-------------|-------------------|-------------|---------|----------|---------|------|
| | | α | β | α | β | |
| 15 | 2G | Mean | 1.00 | 0.80 | 1.00 | 1.16 |
| | | Median | 0.55 | 0.71 | 1.30 | 0.96 |
| | | ME-Median | 0.71 | 0.80 | 1.20 | 1.28 |
| | 3G | Mean | 1.02 | 0.89 | 1.03 | 1.38 |
| | | Median | 0.67 | 0.18 | 1.74 | 1.78 |
| | | ME-Median | 0.72 | 0.47 | 1.78 | 2.49 |
| | Theil | Median | 0.64 | 0.88 | 3.17 | 3.56 |
| | | ME-Median | 0.70 | 0.89 | 3.18 | 3.58 |
| | Siegel | Median | 0.66 | 0.56 | 3.93 | 3.71 |
| ME-Median | | 0.71 | 0.60 | 3.98 | 3.76 | |
| 35 | 2G | Mean | 1.00 | 0.76 | 1.00 | 1.06 |
| | | Median | 0.67 | 0.81 | 1.29 | 1.81 |
| | | ME-Median | 0.74 | 0.88 | 1.32 | 2.03 |
| | 3G | Mean | 1.00 | 0.95 | 1.01 | 0.91 |
| | | Median | 0.58 | 0.70 | 1.33 | 1.31 |
| | | ME-Median | 0.64 | 0.81 | 1.47 | 1.43 |
| | Theil | Median | 0.65 | 0.93 | 3.11 | 1.09 |
| | | ME-Median | 0.65 | 0.93 | 3.18 | 1.22 |
| | Siegel | Median | 0.65 | 0.51 | 2.92 | 1.29 |
| ME-Median | | 0.65 | 0.56 | 3.08 | 1.35 | |
| 95 | 2G | Mean | 1.00 | 0.89 | 1.00 | 1.67 |
| | | Median | 0.82 | 0.44 | 1.15 | 2.48 |
| | | ME-Median | 0.83 | 0.44 | 1.23 | 2.57 |
| | 3G | Mean | 1.00 | 1.06 | 1.00 | 1.99 |
| | | Median | 1.64 | 0.79 | 2.66 | 2.36 |
| | | ME-Median | 0.98 | 0.99 | 2.71 | 2.92 |
| | Theil | Median | 0.97 | 1.64 | 2.82 | 2.20 |
| | | ME-Median | 0.99 | 1.66 | 2.87 | 2.31 |
| | Siegel | Median | 0.98 | 1.03 | 2.50 | 2.53 |
| ME-Median | | 0.99 | 1.57 | 2.61 | 2.60 | |

*2G: Two groups and 3G: Three groups

less than 2.5%. On the other hand and since we believe that the people are the real wealth of a nation, the Human Development Index (HDI) is developed to create an enabling environment for people to enjoy long, healthy and creative lives. The HDI proposed (Sen, 2009; Eneh and Nkamnebe, 2011; UNDP, 1990) as a single statistic to serve as a frame of reference for both social and economic development by combining indicators of life expectancy (X1), literacy (X2) and (the log of) real GDP percapita (X3). The HDI is constructing in three steps:

- **Step 1:** Compute the deprivation indicator for the jth country with respect to the ith variable as:

$$I_{ij} = \frac{\text{Maximum}_j(x_{ij}) - X_{ij}}{\text{Maximum}_j(X_{ij}) - \text{Minimum}_j(X_{ij})};$$

$i = 1, 2, \dots, 3$ and $j = 1, 2, \dots, \text{NOC}$

where, NOC represents the number of countries included in the study

- **Step 2:** Define the average deprivation indicator:

$$I_j = \sum_{i=1}^2 I_{ij}; j = 1, 2, \dots, \text{NOC}$$

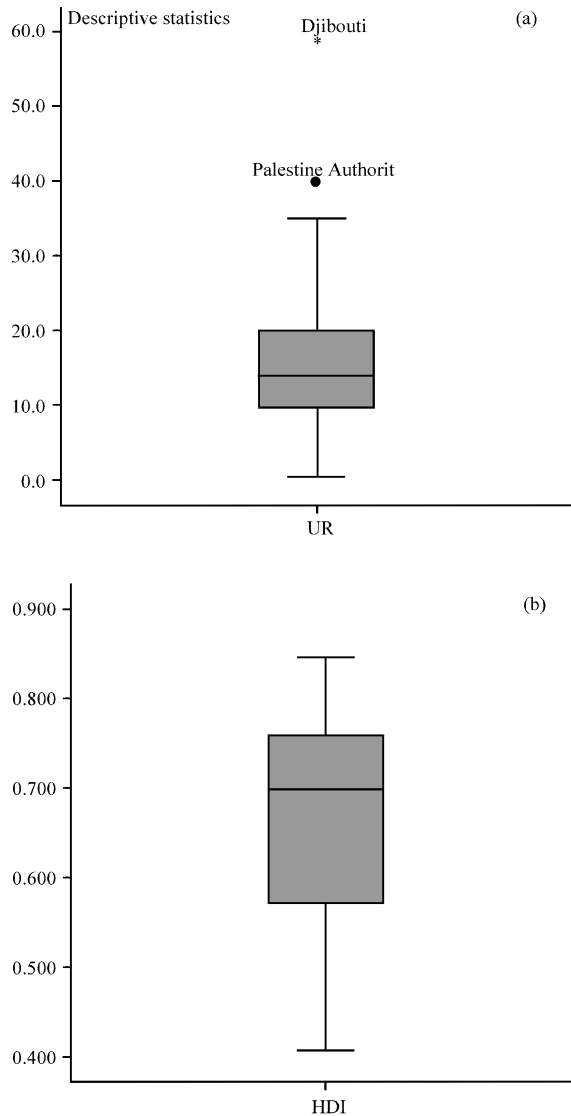


Fig. 2(a-b): Box-plot for UR and HDI in the Arab's States

- **Step 3:** Define the HDI indicator as one minus the deprivation indicator:

$$HDI_j = 1 - I_j; j = 1, 2, \dots, NOC$$

Arab's states data: The data about the twenty two Arab states in Appendix 1 were collected from two formal international databases; CIA world fact book and UNDP organization web page in November, 2011. However, one state has incomplete data (Somalia) and it is excluded from the analysis. Moreover, it could be noted from the box plot given in Fig. 2 that Djibouti and Palestine Authority

Table 3: Quadrants limits of HDI and UR

| Quadrant | HDI ⁽¹⁾ | UR ⁽²⁾ |
|-----------|----------------------------|----------------------------|
| Very high | 0.8 or more | 8.75 or more |
| High | (0.697, 0.795) | (5.25, 8.75) |
| Medium | (0.51, 0.697) | (2.41, 5.25) |
| Low | Less than or equal to 0.51 | Less than or equal to 2.41 |

(1) Classification based on UNDP (2011) report and (2) Classifications based on the 10, 25 and 50th percentiles of the world's UR

have an extreme value of UR; therefore, a robust estimation method is required to avoid the effect of the outliers.

For better data description and countries classifications, we locate the UR and HDI into different quadrants (Table 3).

The classification results in Fig. 3; indicate that there are only two Arab countries with Very high-HDI and Low-UR (UAE and Qatar), on the other hand, Bahrain has strong HDI but it is weak in term of UR. It is worth to say that there are five Arab states (Yemen, Comoros, Sudan, Mauritania and Djibouti) reflect very negative results in both variables HDI and UR.

A general comment, by comparing Arab's states with the world; it could be concluded that this region has a very high UR (in average, 17.571) comparing to the world rate (8.8) and also it has a medium HDI (in average, 0.6503) comparing to the world HDI (0.683). That means many of the economical, educational and social policies should be improved or totally changed by the governments of these states in order to be listed in the world civilized map.

Modeling UR and HDI data: Since a smaller UR more effort in the HDI, then it is expected as showed in the scatter plot (Fig. 4) that there is a linear (negative) relationship between HDI and UR variables. Moreover, different universal resources reports different indicators (with closed values) and some of these indicators were estimated from a historical data or rounded to a three digit values; therefore, the classical regression model is not suitable for these data; instead, the model given in (1)-(2) will be a good formulation of the relationships between the UR (x) and HDI (y). Hence, the suggested compounded URFM for HDI and UR could be formulated as:

$$HDI = \alpha + \beta (UR - \delta) + \epsilon$$

where, α and β are the intercept and the slope, respectively and δ and ϵ are the error terms.

Maximum entropy based estimators: The relationship between HDI and UR is modeled by estimation the

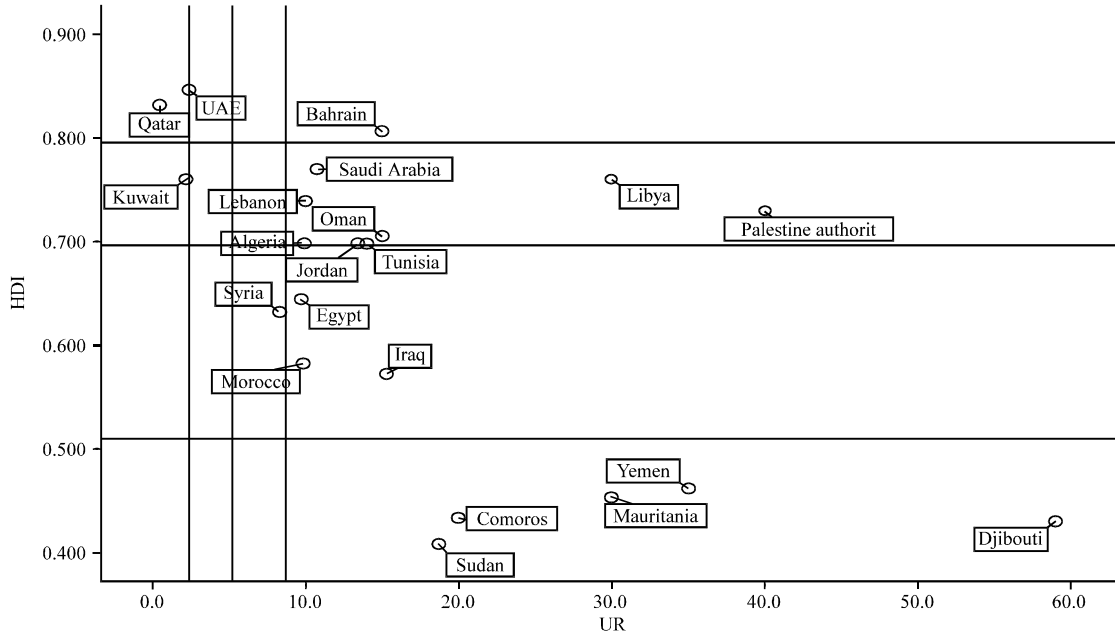


Fig. 3: Arab's State classification with respect to UR and HDI

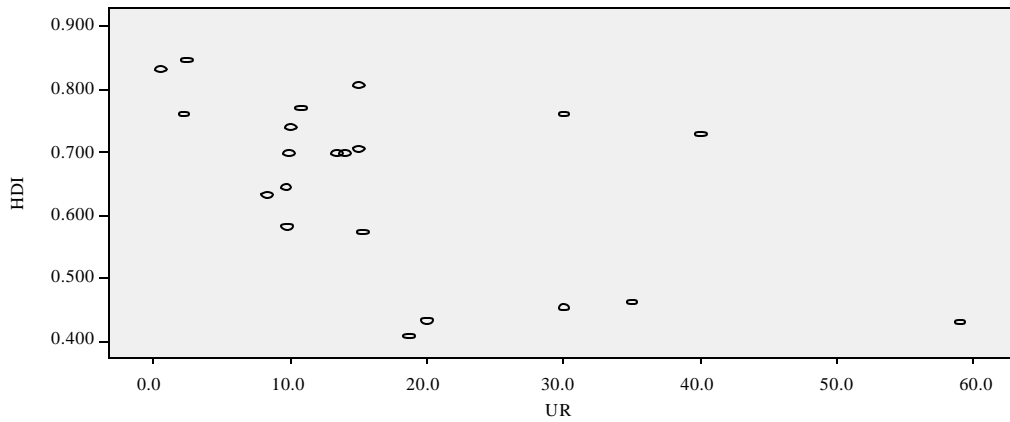


Fig. 4: Scatter diagram showing the relationships between the UR and HDI

Table 4: Estimated parameters of HDI and UR model by the proposed methods

| Method | $\hat{\alpha}$ | $\hat{\beta} \times 10^3$ | Mean sum of squared perpendicular residuals $\times 10^3$ | URFM \hat{R}^2 |
|--------|----------------|---------------------------|---|------------------|
| 2 G | 0.8871 | -17.11 | 0.6871 | 0.4711 |
| 3 G | 0.6442 | -10.467 | 0.5883 | 0.5024 |
| Theil | 0.7739 | -6.4896 | 0.4546 | 0.5059 |
| Siegel | 0.7805 | -7.4389 | 0.3527 | 0.5385 |

unknown parameters with the four modified estimators based on the maximum entropy median. The estimation results along with the estimated coefficient of determination and the mean sum of squared perpendicular residual are given in Table 4.

The results in Table 4 indicate that in terms of mean squared perpendicular residual and in terms of coefficient of determination, the repeated median estimator gave better fit to the data. In general, all models suggested that whenever the country improve the citizen life by decreasing the unemployment rate, the human development index increases. Fig. 5 contains plots of the residuals against the predicted values for HDI and UR. There are no suggestions in any of these plots in order to say that symmetric deviations from the fitted response plan are presented.

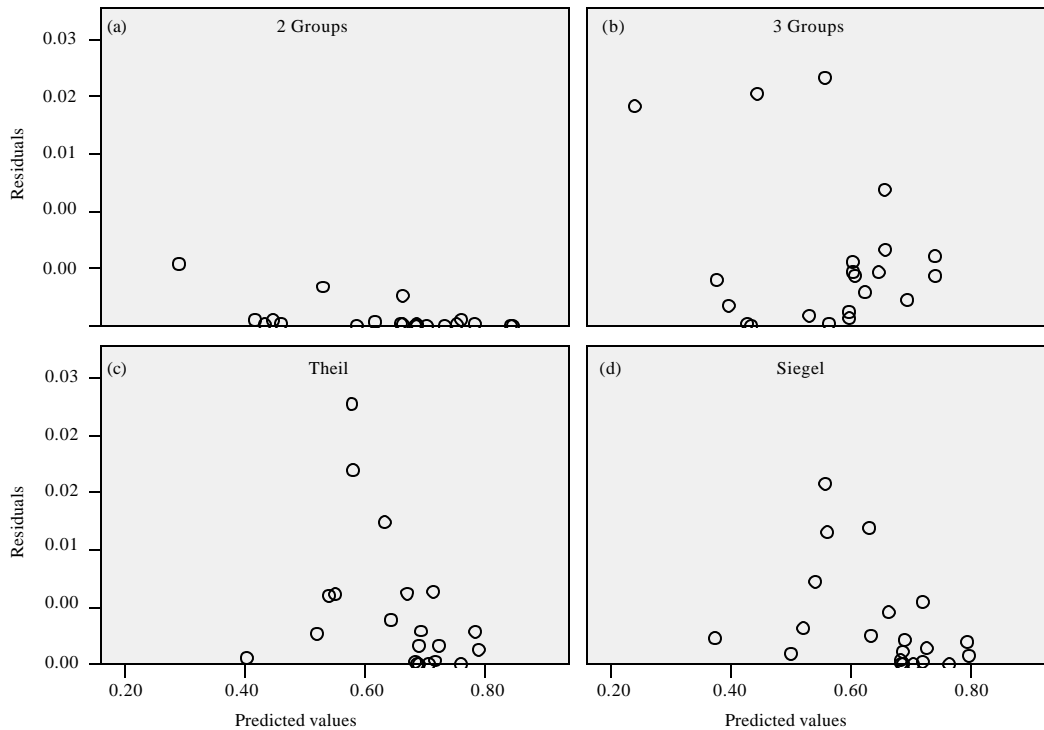


Fig. 5: Predicted values vs Residuals for the proposed estimators

CONCLUSION

This study gives the researcher a more precise method for estimating the parameters of the URFM model when the data contains outliers. The main idea of the proposed estimator is to use the maximum entropy median instead of the sample mean or the sample median in Wald-type and Theil-type estimators. Also, the novel coefficient of determination when both variables are measured with error give a new trend of explanation of the fitted URFM. The Monte Carlo simulations provide a good evidence for the superiority of the proposed model on the classical MLE in terms of the efficiency when the data contains outliers. Also, the real data analysis was used to study the effect of the UR on the HDI based on the proposed estimation method suggested of using the modified repeated median methods. All methods suggested a negative relationship between the unemployment rate in the Arab's states and the human development index. The data analysis suggested that the Arab's states should change their economical, educational and social policies in order to be listed in the world civilized map.

APPENDIX

Appendix 1: HDI and UR data in the Arab states

| Country | Unemployment rate ⁽¹⁾ | HDI ⁽²⁾ |
|------------------------------------|----------------------------------|--------------------|
| Algeria | 9.9 | 0.698 |
| Bahrain | 15.0 | 0.806 |
| Comoros | 20.0 | 0.433 |
| Djibouti | 59.0 | 0.430 |
| Egypt | 9.7 | 0.644 |
| Palestine authority ⁽³⁾ | 40.0 | 0.641 |
| Iraq | 15.3 | 0.573 |
| Jordan | 13.4 | 0.698 |
| Kuwait | 2.2 | 0.760 |
| Lebanon | 10.0 | 0.739 |
| Libya | 30.0 | 0.760 |
| Mauritania | 30.0 | 0.453 |
| Morocco | 9.8 | 0.582 |
| Oman | 15.0 | 0.705 |
| Qatar | 0.5 | 0.831 |
| Saudi Arabia | 10.8 | 0.770 |
| Sudan | 18.7 | 0.408 |
| Somalia | *** | 0.284 |
| Syria | 8.3 | 0.632 |
| Tunisia | 14.0 | 0.698 |
| UAE | 2.4 | 0.846 |
| Yemen | 35.0 | 0.462 |

(1) World fact book; (2) UNDP organization and (3) The statistics are available on Gaza Strip only; *** missing data

REFERENCES

- Al-Nasser, A.D. M.E. Al-Haj and A. Al-Masri, 2005. On the use of l-statistics in wald-type estimators for fitting a straight line when both variables are subject to error. *Pak. J. Stat.*, 21: 179-194.
- Al-Radaideh, A., A. Al-Nasser and E. Ciavolino, 2010. Efficient wald type estimators for simple linear measurement error model. *Asian J. Math. Stat.*, 3: 82-92.
- Bani-Mustafa, A. and A.D. Al-Nasser, 2008. On estimating the variance of maximum entropy median for odd sample size. *Pak. J. Statist.*, 24: 1-10.
- Bartlett, M.S., 1949. Fitting straight line when both variables are subject to error. *Biometrics*, 5: 207-212.
- Carroll, R.J., A. Delaigle and P. Hall, 2009. Nonparametric prediction in measurement error models. *J. Am. Stat. Assoc.*, 104: 993-1003.
- Carroll, R.J., D. Ruppert, C.M. Crainiceanu, T.D. Tosteson and M.R. Karagas, 2004. Nonlinear and nonparametric regression and instrumental variables. *J. Am. Stat. Assoc.*, 99: 736-750.
- Cheng, C.L. and J.W. Van Ness, 1999. *Statistical Regression with Measurement Error*. Arnold, New York, USA.
- Eneh, O.C. and A.D. Nkamnebe, 2011. Gender gap and sustainable human development in Nigeria: Issues and strategic choices. *Asian J. Rural Devel.*, 1: 41-53.
- Fah, C.Y., A.G. Hussin and O.M. Rijal, 2007. An investigation of causation: the unreplicated linear functional relationship model. *J. Applied Sci.*, 7: 20-26.
- Fasoranti, M.M., 2009. The influence of the deregulated telecommunication sector on Urban employment generation in Nigeria. *J. Econ. Theor.*, 3: 57-62.
- Fuller, W.A., 1987. *Measurement Error Models*. Wiley, New York.
- Hussin, A.G., N.R.J. Fieller and E.C. Stillman, 2009. Parameter estimation in the unreplicated functional model based on a grouping algorithm. *Eur. J. Sci. Res.*, 33: 220-233.
- Huwang, L. and J. Yang, 2000. Trimmed estimation in the measurement error model when the covariate has replicated observations. *Proc. Natl. Sci. Counc. Repub. China Part A Phys. Sci. Eng.*, 24: 405-412.
- Jonathan, G. and I. Terence, 2009. Methods of fitting straight lines where both variables are subject to measurement error. *Curr. Clin. Pharmacol.*, 4: 164-171.
- Kayani, Z.I. and A.A. Ghani, 2003. Intercountry inequality in some measure of well-being: 1960-1992. *J. Applied Sci.*, 3: 120-124.
- Kendall, M.G. and A. Stuart, 1979. *The Advanced Theory of Statistics*. 2nd Edn., Griffin, London.
- Maritz, J.S., 1979. On theil's method in distribution-free regression. *Aust. J. Stat.*, 21: 30-35.
- Nair, R.K. and P.M. Shrivastava, 1942. On a simple method of curve fitting. *Sanyukhya*, 6: 121-132.
- O'Driscoll, D. and D.E. Ramirez, 2011. Geometric view of measurement errors. *Commun. Stat. Simulat. Comput.*, 40: 1373-1382.
- Sen, A., 2009. *The Idea of Justice*. The Belknap Press of Harvard University Press, Cambridge, MA, ISBN-13: 978-0674036130, Pages: 496.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's Tau. *J. Am. Stat. Assoc.*, 63: 1379-1389.
- Siegel, A.F., 1982. Robust regression using repeated medians. *Biometrika*, 69: 240-244.
- Solari, M.E., 1969. The "maximum likelihood solution" of the problem of estimating a linear functional relationship. *J. R. Statist. Soc. B*, 31: 372-375.
- Tan, C.H. and A.S. Santhapparaj, 2007. Macroeconomic determinants of skilled labour migration: The case of Malaysia. *J. Applied Sci.*, 7: 3015-3022.
- Theil, H. and P.C. O'Brien, 1980. The median of the maximum entropy distribution. *Econ. Lett.*, 5: 345-347.
- Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. III. *Proc. Koninklijke Nederlandse Akademie van Wetenschappen A*, 53: 1397-1412.
- UNDP, 1990. *Human Development Report 1990*. Oxford University Press, New York.
- UNDP, 2011. *Human development index and its components*. United Nations Development Program, pp: 127-131.
- Wald, A., 1940. The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.*, 11: 284-300.