Data envelopment analysis based robust regression for assessing healthcare systems of eastern mediterranean

**Ali Sadiq Mohommed Bager**
Al-Muthanna University, Samawa, Iraq
Email: ali.sadig@mu.edu.iq

**Hajem A. Daham**
Department of Mathematics, College of Education for Pure Science, Al-Muthanna University, Samawa, Iraq
Email: hajem.daham@mu.edu.iq

**Kadhem Al-Daffaie**
Department of Banking and Financial Sciences, College of Administration and economics, Al-Muthanna University, Samawa, Iraq
*Corresponding author email: kadhem.aldaffaie@mu.edu.iq

**Abstract**---Healthcare systems efficiency is a ratio of service delivery system outputs to inputs to achieve optimal strategy for decision makers. In this paper, we investigated the efficiency of Eastern Mediterranean health observatory systems. For this purpose, a data envelopment analysis model formulated to evaluate the efficiency of healthcare systems response determinants for the Eastern Mediterranean countries. To find the important inputs and outputs that effect the efficiency, a robust regression model estimated, and the result explained which of inputs and outputs have an effect of the efficiency of these systems.

**Keywords**---DEA, robust regression, health care systems, eastern mediterranean.

**Introduction**
Timely and robust information and data in healthcare sector, indeed, properly health management, evidence based on decisions making, reliable allocation of resources, controlling and also active and continuous evaluation of the healthcare systems situation. All these mentioned activities are essential for developing policies of the healthcare sector. Although, the demand for healthcare data and information is recently increased, responses to these requirements have been
obstructed because of scattering, gap and lacking strength of international or even local healthcare data and information systems. One of the global organizations is World Health Organization (WHO) who take responsibility of providing technical supporting to countries members in this organization. The WHO undertaken collecting a strength and robust healthcare information of various countries and regions systems around the world. This allows accessing timely and robust information and data as a clear evidence to evaluate systems responses of healthcare sector. Healthcare system efficiency is a comparison of outputs of the system, such as healthcare facilities and outcomes, with inputs like cost represented by expenditure offered by government. Efficiency in this case is a ratio of outputs to inputs using Data Envelopment Analysis (DEA) or other frontier analysis techniques, which has been used widely to assess efficiency of healthcare systems. Pelone et al [18], Kohel et al. [14] and recently Zakowska et al [23] reviewed references related to the use of DEA in healthcare sectors. For instance, Amado and Dyson [3] applied DEA in healthcare for performance evaluation and improvement in primary health care systems. Some references have considered the application of DEA for specific countries, for instance, Masiye [16] investigated the performance of Zambian hospitals using DEA methodology. While, Carrillo et al [6] established a study to evaluated the efficiency of systems of healthcare areas in Spain and find that areas using healthcare inputs are more efficiently rather than outputs. Ibrahim and Daneshvar [12] evaluated the healthcare system in Lebanon using an updated by DEA model base on some inputs and output variables.

On the other hand, Many references have considered the evaluation and assessment of health care of different regions by the use of DEA, for example, Afonso and Aubyn [1] estimated a semi-parametric model based regressing data envelopment analysis for assessing OECD countries health care. Hadad et al., [10] also applied the DEA for healthcare systems based on the determinants associated with efficiency of the healthcare systems of the OECD region. Later, Behr and Theune, [5] investigated the efficiency of OECD health data based on the focus of different healthcare systems aspects. Asandului et al [4] also applied a DEA approach for measuring the healthcare systems in Europe. Similarly, for Euorp countries, Gómez-Gallego et al [9] evaluated the efficiency of health care systems based on Fuzzy Data Envelopment Analysis approaches. However, Zari et al [24] proposed a hybrid DEA based game theory to measure the performance of group of regional health care centres. While Top et al [20] measured the efficiency of number of African countries applying DEA approach. For Asian countries, Ahmed et al [2], have also applied the DEA based Censored Tobit regression and smoothed bootstrap approach to measure the efficiency of health care systems of Asian countries. Recently, Jordi et al. [13] estimated the national health efficiency of national health. A two-stage efficiency analysis based DEA investigated how efficiently of health care systems for number of countries around the world. Other references are also conducted to assess the health care systems in other different areas around the world considering different inputs and outputs [11, 17, 18, 22].

In this paper, we investigated the efficiency of Eastern Mediterranean health observatory. A data Envelopment Analysis (DEA) model formulated to estimate the efficiency of health care systems response determinants of the Eastern
Mediterranean countries. To analyse the important inputs and outputs that affect the efficiency of health care systems in those countries, a robust regression model is applied.

**Data Envelopment Analysis (DEA)**

Charnes, Cooper and Rhodes [7] first developed Data Envelopment Analysis in 1978. It is assumed that a number of decision making units (nDMU: DMU₁, DMU₂, ..., DMUₙ, where j=1,2,...,n) with a number of m inputs, \( x_{ij} \) \( i=1,2,...,m \), and number of p outputs, \( y_{ij} \) \( l=1,2,...,p \). Let the weight of inputs is \( v_i \) \( i=1,2,...,m \) and the weight of outputs is \( u_l \) \( l=1,2,...,p \), then:

\[
\begin{align*}
\min h &= \frac{\sum_{j=1}^{n} u_l y_{lj}}{\sum_{i=1}^{m} v_i x_{ij}} & \quad \text{......(1)} \\
\text{s.t.} & & \\
\frac{\sum_{j=1}^{n} u_l y_{lj}}{\sum_{i=1}^{m} v_i x_{ij}} & \leq 1 & \forall j \\
u_l, v_i & \geq 0
\end{align*}
\]

Note that the above DEA model can be convert to linear programming model, which can solved by simplex or any other solution methods for linear programming. The efficiency \( e_j \) of each decision making unit can be found by solving the linear program model for the DEA as follows:

\[
\begin{align*}
\max e_j &= \sum_{j=1}^{n} u_l y_{lj} & \quad \text{......(2)} \\
\text{s.t.} & & \\
\sum_{i=1}^{m} v_i x_{ij} &= 1 \\
\sum_{j=1}^{n} u_l y_{lj} - \sum_{i=1}^{m} v_i x_{ij} & \leq 0 & \forall j \\
u_l, v_i & \geq 0
\end{align*}
\]

**Numerical Experiment**

In this section, in order to assess the health systems responses in the Eastern Mediterranean (EM) countries. Data is collected from the World Health Organization website https://rho.emro.who.int/. A number of 17 of the EM countries are considered and data are collected for years 2005-2018. The inputs and outputs selected based on the importance of each variable (input or output) as in Table1. The first input \( (x_1) \) is the Domestic General Government Health Expenditure, which is the budget that each country is allocated to the health care system. Input 2 represents the Hospital beds per 10 000 population \( (x_2) \), while the third input \( (x_3) \) is the health workforce which is consisted of the number Physicians, Nursing and midwifery, Dentists and Pharmacists per 10 000 population for each country. The fourth input \( (x_4) \) represents the population size for each country. Outputs on the other hand are the Life expectancy at birth-Output 1 \( (y_1) \), Out of Pocket (OOP) of Current Health Expenditure-Output 2 \( (y_2) \) and the Primary health care facilities per 10 000 population – Output 3 \( (y_3) \).
Table 1
Inputs and Outputs

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic General Government Health Expenditure - Input 1 ((x_1))</td>
<td>Life expectancy at birth - Output 1 ((y_1))</td>
</tr>
<tr>
<td>Hospital beds per 10 000 population - Input 2 ((x_2))</td>
<td>Out of Pocket of Current Health Expenditure - Output 2 ((y_2))</td>
</tr>
<tr>
<td>Health Workforce - Input 3 ((x_3))</td>
<td>The primary healthcare facilities per 10 000 population - Output 3 ((y_3))</td>
</tr>
<tr>
<td>Population size (in thousands) - Input 4 ((x_4))</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 1. Efficiency Scores of DEA model

The DEA mathematical model is coded in Matlab. Results in Figure 1 shows that some EM countries such as Saudi Arabi, Qatar, Pakistan, Palestine, Morocco, Iraq and Afghanistan are have efficient health systems responses based on the four and three used inputs and outputs through the years 2005-2018. However, other countries have lack of efficiency for some years. For instance, Bahrain has lack of efficiency for the years 2005, 2006 and 2007, while it becomes efficient for the following period. Similarly, the United Arab Emirates has un efficient health system responses for the years 2010 and 2011, while it has efficiency in the period before and after 2010 and 2011. On the other hand, we can see from Table 2 that the health system response in Sudan is inefficient all along the last 14 years.

Robust regression model

The classical methods in estimating the parameters of the model are inaccurately the data when there is an effect in one of the regression assumption. The
presence of outliers and the random error distribution is not normal distribution that suits the method adopted in the estimation. If one or more abnormal values are found, it will lead to a defect in the characteristics of least squares estimators and that the robust estimator is the one who maintains the desired properties of the estimators when some of the model hypotheses are violated Li [15]. Wagner and Shimshak [21] have applied the stepwise selection of variables in data envelopment analysis. Robust regression approach can have used efficiently to investigate the effect of variables, Susanti and Pratiwi [19].

**Robust estimation method**

The estimation maximum likelihood and M method is robust estimation, then the M estimation as Gogoi and Gogoi [8]:

\[
\hat{b} = (\tilde{b}_n)_{x_1, x_2, \ldots, x_3} = b \quad \ldots \ldots (1)
\]

Then

\[
E\{\hat{b} = (\tilde{b}_n)_{x_1, x_2, \ldots, x_3}\} = b
\]

For eq.(1) we will get the estimator:

\[
\hat{b}_M = \min v(y_i - \sum_{i=0}^{z} x_i b_z) \ldots \ldots \ldots (2)
\]

We have to solved the above equation (2)

\[
\min \sum_{i=1}^{n} u(\varepsilon_i) = \min \sum_{i=1}^{n} u(u_i * \frac{1}{\sigma}) = \min \sum_{i=1}^{n} u(\frac{1}{\sigma} * (y_i - \sum_{i=0}^{z} x_i b_z))
\]

The obtain for equation (2) and we estimator for \(\sigma\)

\[
\hat{\sigma} = MAD \left(\frac{1}{0.645}\right) = \left(\frac{1}{0.645}\right) (median|\varepsilon_i - median(\varepsilon_i)|)
\]

For the function of \(v\) we use Tukeys bisquare objective function

\[
v(\varepsilon_i) = \begin{cases} 
\frac{\varepsilon_i^2}{2} - \frac{\varepsilon_i^4}{2\sigma^2} + \frac{\varepsilon_i^6}{6\sigma^4} & |\varepsilon_i| \leq \sigma \\
\frac{\varepsilon_i^2}{6} & |\varepsilon_i| > \sigma 
\end{cases}
\]

Furthermore then see for first partial derivative \(\hat{b}_M\) for \(b\) so that

\[
\sum_{i=1}^{n} x_{ij} \omega (\frac{1}{\sigma} * (y_i - \sum_{i=0}^{z} x_i b_z)) = 0 \quad j = 1, 2, \ldots, z \ldots (3)
\]

Then the \(\omega = v\) and \(x_{ij}\) is the observation on the independent variable \(\sum_{i=1}^{n} x_{ij} = 1\)

Then the equation 3 solution became as:

\[
w(\varepsilon_i) = \omega (\frac{1}{\sigma} * (y_i - \sum_{i=0}^{z} x_i b_z)) \\
(\frac{1}{\sigma} * (y_i - \sum_{i=0}^{z} x_i b_z))
\]

The \(\varepsilon_i = \frac{\varepsilon_i}{\sigma}\) also can write as the following form:
\[ u(\varepsilon_i) = \begin{cases} 
(1 - (\varepsilon_i^2)^2) & |\varepsilon_i| \leq c \\
0 & |\varepsilon_i| > c 
\end{cases} \]

\[ \hat{x}_i \hat{w}_i x b = x \hat{W}_i y \]

Then the \( w_i \) is matrix \( n \times n \) with the diagonal element or are weight.

### Table 2
Descriptive Statistics

<table>
<thead>
<tr>
<th>V</th>
<th>Count</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>235</td>
<td>7.190301</td>
<td>3.141636</td>
<td>0.866</td>
<td>22.6</td>
</tr>
<tr>
<td>x_2</td>
<td>235</td>
<td>15.83825</td>
<td>7.965885</td>
<td>0.01</td>
<td>80.1</td>
</tr>
<tr>
<td>x_3</td>
<td>235</td>
<td>13.96679</td>
<td>8.086802</td>
<td>1.3</td>
<td>54.58</td>
</tr>
<tr>
<td>x_4</td>
<td>235</td>
<td>51.49323</td>
<td>111.8912</td>
<td>1.04</td>
<td>838</td>
</tr>
<tr>
<td>y</td>
<td>235</td>
<td>37.69902</td>
<td>5.989972</td>
<td>27.77</td>
<td>61.39</td>
</tr>
</tbody>
</table>

The value of \( c = 4.685 \) Tukeys bisquare weighted so equation (3)
\[ \sum_{i=1}^{n} X_{ij} w_i (y_i - \sum_{j=1}^{x} X_{ij} b) \]

As shown in Table 2, the descriptive statistics of variables of the DEA model can be explained as follows: mean of variable \( x_1 \) of the sample is approximated to 17.19 %, with a standard deviation of 3.14, while the mean value for the \( x_2 \) variable is 15.83% with a standard deviation of 7.96, mean of variable \( x_3 \) is 13.96% with a standard deviation of 7.96 and the result also shows that the mean of the \( x_4 \) variable is 51.49% with a standard deviation of 111.89, and finally the mean of \( y \) variable is 37.69 with a standard deviation of 5.9%.

On the other hand, when we solve the robust estimation method as shown in Table 3, which explains the relationship between the \( y \) variable and the four explanatory variables. We can see from Table 3 that the relations of the dependent variable \( y \) with the three explanatory variables \( x_1, x_2 \) and \( x_4 \) are statistically significant, which refers that these three variables have a direct effect on the variable \( y \). This means that the relationship between the \( x_1 \) and the \( y \) is positive, and the P-value of regression coefficient is 0.000, which refers that any increasing in the \( x_1 \) variable will lead to a corresponding increase in variable \( y \). The P-value of the regression coefficient for variable \( x_2 \) is 0.0002, which means that the relationship between \( x_2 \) and \( y \) variables is also positive so that any increasing in the \( x_2 \) variable will cause a corresponding increasing in \( y \). Similarly, the relationship between the \( x_4 \) and \( y \) variables is also positive with a P-value of the regression coefficient of 0.0001, referring that an increasing in the \( x_4 \) variable responses with an increase of the \( y \) variable.
### Table 3
Result of Robust estimation

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-Statistic</th>
<th>P-Value Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁</td>
<td>2.81818</td>
<td>0.286255</td>
<td>9.845</td>
<td>0.0000</td>
</tr>
<tr>
<td>x₂</td>
<td>0.5601971</td>
<td>0.1453394</td>
<td>3.854</td>
<td>0.0002</td>
</tr>
<tr>
<td>x₃</td>
<td>0.07319678</td>
<td>0.1375531</td>
<td>0.532</td>
<td>0.5951</td>
</tr>
<tr>
<td>x₄</td>
<td>0.0362789</td>
<td>0.0088829</td>
<td>4.084</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

### Conclusion

In this paper, the efficiency of Eastern Mediterranean healthcare systems is investigated. A data envelopment analysis model applied to estimate the efficiency of healthcare systems response. Results show that some countries have full efficiencies for their healthcare systems based on the considered inputs and outputs. To find which inputs and outputs effected the efficiency, a robust regression model is applied. Results also showed that domestic general government health expenditure and population size variables have the most effect on efficiency of healthcare systems.

### References


