International Journal of Mathematics and Statistics Studies

Vol.3, No.3, pp.15-20, May 2015

Published by European Centre for Research Training and Development UK (www.eajournals.org)

# HEALTH CARE EXPENDITURE IN AFRICA – AN APPLICATION OF SHRINKAGE METHODS

**Emmanuel Thompson**
Department of Mathematics, Southeast Missouri State University, One University Plaza, Cape Girardeau, Missouri 63701, USA

**ABSTRACT:** *In the literature, apart from Gross Domestic Product (GDP) per Capita, a number of factors have been identified as non-income determinants of health care expenditure. Most studies linking health care expenditure to income and other important variables have originated in advanced countries. This study considers a linear regression model via shrinkage methods to identify key predictors of health care expenditure in Africa. Shrinkage is a process of estimation where a subset of redundant predictor variables is discarded, leaving only important ones in the linear regression model. The study was based on 42 African countries for the year 2012 and GDP per Capita, Consumer Price Index (CPI), Exchange Rate (EXC), Corruption Perception Index (COP), and Population Density (POP) were identified as key predictors of Health Expenditure per Capita (HEC). A major finding identified the Elastic net model as critical in accurately estimating HEC in Africa.*

**Keywords**: Least Absolute Shrinkage and Selection Operator, Elastic net, Multicollinearity, Heath Expenditure per Capita, Coordinate Descent

## INTRODUCTION

Income has been identified as a key predictor of health care expenditure, however, there is no unanimity on which other variables may be related to the remaining largely unexplained variation in HEC (see [1] and [2]). In the literature, a number of factors have been identified as non-income determinants of health care expenditure (see [2] and [3]). It is important to place on record that, most studies linking health care spending to income and other important variables have originated in advanced countries. Since the work of [4] and [5] and recently [6], not much is known about Africa. This paper seeks to contribute to filling this gap by identifying key predictors and eventually proposing a predictive model for health care expenditure for Africa. Using 42 African countries for the year 2012, the link between health care expenditure and 10 predictors (see material and method section) is investigated using shrinkage methods. There are occasions when the traditional statistical estimation methods such as the least squares tend to produce imprecise estimates of parameters causing erroneous inferences in the presence of multicollinearity (see [7]). In such situations, it is often advantageous to resort to shrinkage techniques. Shrinkage is a process of estimation where a subset of redundant predictors is discarded, leaving only important ones in the linear regression model. The shrinkage methods work by imposing a penalty on the least squares estimation method. Various assumptions have been made in the literature where the penalty of $l_1 - norm,\ l_2 - norm$ or both which represents the penalty parameter $(\lambda)$ were used to influence the parameter estimates in order to minimize the impact of collinearity. This study focuses on two of the shrinkage methods- LASSO and Elastic net. LASSO, which stands for Least Absolute Shrinkage and Selection Operator (LASSO), was proposed by [8] and has the property of shrinking some of the regression coefficients to exactly zero. The Elastic net technique on the other hand, was introduced by [9] as an improved version of the LASSO. The method selects predictors like

LASSO and shrinks the coefficients of correlated predictors similar to ridge regression. While previous studies focused much on time series and panel data models, the current study emphasizes supervised learning and predictive modelling. The rest of the paper is planned as follows. Section 2 describes the material and the method. Section 3 gives the empirical results. Section 4 provides the concluding remarks.

## MATERIAL AND METHOD

This study utilized data from Africa Economic Outlook (see [10]) and the World Bank (see [11]) open data websites. The variables used were as follows:

**Table 1. Definition of variables and sources**

| Variable | Definition | Source |
|---|---|---|
| HEC | Health Expenditure per Capita, PPP (constant 2005 international $) | The World Bank |
| GDP | GDP per Capita, PPP (constant 2011 international $)<br>PPP stands for purchasing power parity | The World Bank |
| CPI | Consumer Price Index (2010 = 100) | The World Bank |
| INF | Inflation Rate (%) | Africa Economic Outlook |
| EXC | Exchange Rate(LCU / $ )<br>LCU stands for local currency unit | Africa Economic Outlook |
| BMN | Broad Money (LCU billion )<br>Note: LCU stands for local currency unit | Africa Economic Outlook |
| UEM | Unemployment Rate (%) | Africa Economic Outlook |
| COP | Corruption Perception Index | Africa Economic Outlook |
| POP | Population Density (people per sq. km of land area) | The World Bank |
| LEB | Life Expectancy at Birth, Total (years) | The World Bank |
| TUB | Tuberculosis (New and relapse cases) | Africa Economic Outlook |

HEC was used as the response variable to capture health care expenditure, while the rest were predictors of HEC. The data structure was a cross section of 53 African countries during 2012. Some countries were excluded from the study due to missing data and the possibility to obtain them proved futile. All variables were transformed into logarithms prior to analyses. Data were analyzed using descriptive and inferential statistics. The Pearson correlation coefficient was used to describe the strength of the linear relationship between the response variable and each of the predictors and also among the predictors as well. LASSO and Elastic net estimation methods were used to identify important predictors of HEC. Prior to estimating regression coefficients, the data set was partitioned into a training set (25 observations) and a testing set (17 observations). The training set was used to estimate the LASSO and Elastic net regression models. The testing set was then used to check the performance of the models based on the training set. The mean square error (MSE) was used as a metric to assess the error of prediction of the models. The optimal $\lambda$ for both LASSO and Elastic net were estimated by the cross-validation (CV) method with MSE criterion as suggested by [12].



ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

**Statistical Model**

Consider the general linear model of the Gaussian family where i represents countries.

$$y_i = \beta^T x_i + \epsilon_i,$$

where $(x_i, y_i);\ i = 1, 2, \dots, N$ are a sample of N independent and identically distributed (i.i.d) random vectors, where $x_i = (x_{i1}, x_{i2}, \dots, x_{ik}) \in \mathcal{R}^k$ is the random vector of observations about k predictors for the $i^{th}$ sample unit and $y_i \in \mathcal{R}$ is the corresponding response vector.
The vector of $(k+1)$ estimates $(\hat{\beta}_0, \hat{\beta})$ of regression coefficients were obtained by applying the coordinate descent (see [12]) to solve the optimization problem whose objective function is given by

$$\min_{(\beta_0,\beta)\in\mathcal{R}^{k+1}} \sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^T\beta\right)^2 \lambda[(1-\alpha)\,\|\beta\|_2^2/2 + \alpha\|\beta\|_1^2]$$

where $\lambda \geq 0$ and $0 \leq \alpha \leq 1$. The ridge regression coefficients are obtained by setting $\alpha = 0$ which is not a subject of consideration in this study. When $\alpha = 1$, the optimized problem produces the LASSO regression coefficients and $0 < \alpha < 1$ results in the Elastic net regression coefficients.

## RESULTS

Table 2 shows the strength of the linear relationship among the variables. From the table, it is evident that some of the predictors were correlated. High and positive linear correlation were observed between CPI and INF (0.917) and somewhat moderately high and positive linear correlation between GDP and LEB (0.527).

**Table 2. Pearson Correlation Coefficients**

| | HEC | GDP | CPI | INF | EXC | BMN | UEM | COP | POP | LEB | TUB |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **HEC** | 1.000 | 0.812 | -0.134 | -0.105 | -0.268 | -0.137 | 0.514 | 0.566 | 0.136 | 0.306 | 0.302 |
| **GDP** | 0.812 | 1.000 | -0.126 | -0.136 | -0.288 | -0.100 | 0.355 | 0.329 | 0.076 | 0.527 | 0.067 |
| **CPI** | -0.134 | -0.126 | 1.000 | 0.917 | 0.332 | 0.330 | -0.100 | -0.383 | 0.076 | -0.106 | 0.125 |
| **INF** | -0.105 | -0.136 | 0.917 | 1.000 | 0.227 | 0.269 | -0.023 | -0.320 | 0.117 | -0.093 | 0.024 |
| **EXC** | -0.268 | -0.288 | 0.332 | 0.227 | 1.000 | 0.351 | -0.286 | -0.202 | -0.028 | -0.207 | -0.036 |



ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **BMN** | -0.137 | -0.100 | 0.330 | 0.269 | 0.351 | 1.000 | -0.308 | -0.213 | 0.041 | -0.065 | 0.238 |
| **UEM** | 0.514 | 0.355 | -0.100 | -0.023 | -0.286 | -0.308 | 1.000 | 0.257 | -0.172 | -0.004 | 0.209 |
| **COP** | 0.566 | 0.329 | -0.383 | -0.320 | -0.202 | -0.213 | 0.257 | 1.000 | 0.222 | 0.245 | -0.028 |
| **POP** | 0.136 | 0.076 | 0.076 | 0.117 | -0.028 | 0.041 | -0.172 | 0.222 | 1.000 | 0.192 | -0.077 |
| **LEB** | 0.306 | 0.527 | -0.106 | -0.093 | -0.207 | -0.065 | -0.004 | 0.245 | 0.192 | 1.000 | -0.155 |
| **TUB** | 0.302 | 0.067 | 0.125 | 0.024 | -0.036 | 0.238 | 0.209 | -0.028 | -0.077 | -0.155 | 1.000 |

Figure 1 shows the paths of estimated coefficients for each of the predictors against log $\lambda$ using Elastic net (left panel) and the LASSO (right panel) algorithms. The figure indicates how the coefficients entered the models as $\lambda$ changes. For small values of $\lambda$, the estimated coefficients were close to that of the least squares. Judging from the figure, it was difficult to identify appropriate value of $\lambda$ in order to select the optimal models from both the LASSO and Elastic net estimation methods. Note that the value of $\alpha$ was set at 0.5 for simplicity in the Elastic net model.

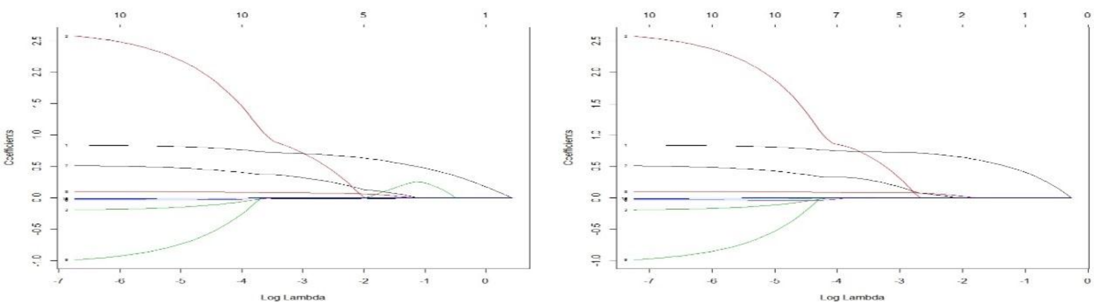

**Figure 1. Estimated Coefficient Paths**

Figure 2 depicts the CV curves (Elastic-net on the left panel and LASSO on the right panel) shown in red dotted lines with normal standard error bands around it. The value of $\lambda$ that minimizes the MSE was 0.047 for the LASSO and 0.104 for the Elastic net.

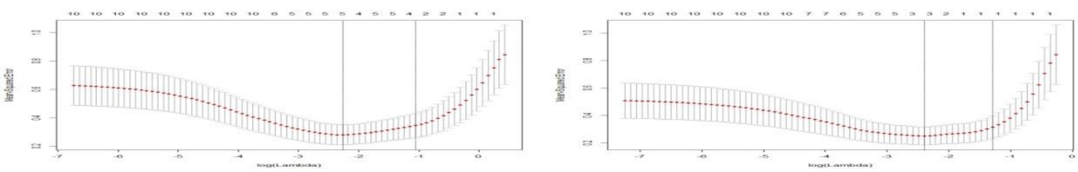

**Figure 2. Plots of MSE against log $\lambda$**


ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Table 3 shows the non-zero estimated coefficients extracted at the optimal value of $\lambda$ for both the LASSO and Elastic net. The non-zero coefficients correspond to predictors: ln (GDP), ln (CPI), ln (EXC), ln (COP), and ln (POP) for both the LASSO and the Elastic net. The MSEs on the basis of the testing set were 0.427 and 0.389 for the LASSO and the Elastic net respectively. A comparison of the MSEs confirms that the Elastic net model provided a better fit.

**Table 3. Non-Zero Estimated Coefficients**

| Coefficient | LASSO | Elastic net |
|---|---|---|
| Intercept | -3.110 | -1.960 |
| ln (GDP) | 0.726 | 0.660 |
| ln (CPI) | 0.353 | 0.243 |
| ln (INF) | - | - |
| ln (EXC) | -0.002 | -0.016 |
| ln (BMN) | - | - |
| ln (UEM) | - | - |
| ln (COP) | 0.187 | 0.196 |
| ln (POP) | 0.080 | 0.069 |
| ln (LEB) | - | - |
| ln (TUB) | - | - |

## CONCLUDING REMARKS

The results of the study suggest the appropriateness of shrinkage methods as a potentially effective analytical tool for identifying important predictors of HEC in the presence of multicollinearity. The study revealed that GDP, CPI, EXC, COP, and POP were the key predictors associated with the non-zero estimated coefficients for both the LASSO and the Elastic net models. A major finding of the study was that it identified the Elastic net model as critical in accurately estimating the predictors of HEC in Africa. The findings of the present study provide a reference for policymakers and governments in Africa in terms of the key predictors of health care expenditure and hope they will give them a consideration in the determination of health care cost. Also, the findings herein have implications for future research. First, the current study did not take into consideration the stage of development of various countries. The inclusion of categories of countries by Gross National Income (GNI) to identify key predictors of HEC is justified. Second, Sparse Principal Component Analysis (SPCA), a modified version of Principal Component Analysis (PCA) produces a more parsimonious model. Future research needs to compare the Elastic net to the SPCA.

## REFERENCES

[1] Wilson, R. M. (1999). Medical care expenditures and gdp growth in OECD nations. American Association of Behavioral and Social Sciences Journal 2, 159-171.
[2] Baltagi, B. H., and Moscone, F. (2010). Health care expenditure and income in the OECD reconsidered: evidence from panel data. Economic Modelling, 27, 804-811.
[3] Culyer, A. J. (1988). Health care expenditures in Canada: Myth and reality; past and future. Tax Paper n. 82, Canadian Tax Foundation.

[4] Gbesemete, K. P. and Gerdtham, U. G. (1992). Determinants of health care expenditure in Africa: a cross-sectional study. World Development, 20, 303-308.


ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

[5] Jaunky, V. C. and Khadaroo, A. J. (2008). Health care expenditure and GDP: an African perspective. Applied Econometrics and International Development, vol. 8, 131-146.

[6] Lv, Z., and Zhu, H. (2014). Health care expenditure and GDP in African countries: Evidence from semi-parametric estimation with panel data. The Scientific World Journal, 1-6.

[7] Muhammad, I., Maria, J., and Muhammad, A. R. (2013). Comparison of shrinkable regression for remedy of multicollinearity problem. Middle-East Journal of Scientific Research, 14, 570-579.

[8] Tibshirani, R. J. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, Series B. 58, 267-288.

[9] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B 67, 301 – 320 67, 301 – 320.

[10] Africa Economic Outlook (2012). Inflation rate, exchange rate, broad money, unemployment rate, corruption perception index, and tuberculosis (new and relapse cases), [Data file]. Retrieved from http://www.africaneconomicoutlook.org/en/statistics/.

[11] The World Bank, World Development Indicators (2012). Health expenditure per capita, GDP per capita, consumer price index, population density, and life expectancy at birth, [Data file]. Retrieved from http://data.worldbank.org/topic

[12] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear model via coordinate descent, Journal of Statistical Software, 33, 1-22.



ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)