International Journal of Management, IT & Engineering
Vol. 9 Issue 1, January 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell’s Directories of Publishing Opportunities, U.S.A

EXPLORATION OF VARIABLES WITH RIDGE REGRESSION IN DATA ENVELOPMENT ANALYSIS

Reshampal Kaur[1]

Dr. H.S. Bhatti[2]

Dr. Monika Aggarwal[3]

Abstract: Tomeasurethe efficiency levels, literature has been advocating the use of Data envelopment analysis as a powerful service management and benchmarking technique. DEA measures the relative efficiency amongst decision making units (DMUs), considering all input and output resources used by them and identifies the most efficient units. Although DEA calculates sources of inefficiency for less efficient units, but it does not provide information related to determinants of efficiency. Particularly, in a multiperiod environment, DEA has to be applied in combination with other techniques for better interpretation of results. This paper attempts to estimate efficiency using DEA in a multiperiod environment in combination with Ridge Regression, using R software with the objective of identifying determinants of efficiency, using database of Indian public sector banks (PSBs) from the year 1998 till 2013.

It was found that only 51% of the PSBs were efficient, ‘Borrowings’ and ‘Deposits’ were found to be main source of inefficiency. ‘Wage bills’ were found to be the main determinant of efficiency. It was concluded that to deal with multiperiod and further multicollinearity in data, the use of DEA in association with Ridge Regression provides better interpretation of results.

Key words: *Efficiency, Data Envelopment Analysis (DEA), Public Sector Banks (PSBs), Decision Making Units (DMUs), Ridge Regression.*

[1] **Research Scholar, Panjab Technical University, Jalandhar, India,**

[2] **HOD & Professor, BBSB Engineering College, Fatehgarh Sahib, India,**

[3] **Associate Professor, UIAMS, Panjab University, Chandigarh, India,**

I. INTRODUCTION

Data envelopment analysis (DEA) is a linear programming-based technique of operations research, which is globally accepted and a preferred tool to measure relative efficiency of peer entities, called decision making units (DMUs). It identifies the best performing units as efficient DMUs and other less efficient units as inefficient DMUs, with an insight of sources of inefficiency and benchmarks for each of the less efficient DMU. Although DEA provides measures of efficiency in a comparative analysis amongst DMUs, in a single time period, but it does not give any information on the determinants of efficiency. In a multiperiod environment, DEA has to be used in combination with other techniques to find results based on overall time period. For this purpose,DEA and regression analysis are the techniques which are quite compatible to be used in combination, for better interpretations of results and to identify the determinants of efficiency. The DEA focuses on the outliers of the datasetto evaluate efficiency levels of DMUs in individual time periods andRegression analysis considers the average performance of all the DMUs in various time periods of a multiperiod study, to establish a relationship between a dependent variable and one or more independent variables by providing a regression equation.

Review of existing literature shows that particularly in Banking sector, researchers have been usingdifferent types of regression analysis for regressing DEA efficiency scores,on some bank specific or environmental variables, other than input- output variables used for DEA.Ariff M. & Can L., (2008), Das A. & Ghosh S., (2009), Sufian F. (2009), Batir T.E., Volkman D.A. &Gungor B., (2017), used DEA followed by Tobit Regression.Chiu Y.H. & Chen Y.C., (2009), used DEAfollowed by stochastic frontier regression.Assaf A.G., Barros C.P. &Matousek R., (2011),Chortareas G.E., Girardone C. and Ventouri A., (2012),Chortareas G.E., Girardone C. and Ventouri A., (2013),Wanke P., Barros C.P. and Emorouznejad A. (2016)used DEA and truncated regression model. Sufian F. (2011a), Chortareas G.E., Garza-Garcia J.G. &Girardone C., (2012),Sufian F. & Habibullah M.S., (2012),Kamarudin F., Sufian F. &Nassir A.M., (2016) employed DEA followed by the panel regression analysis based on the Ordinary Least Square (OLS).Sufian F., (2011b), employed DEA followed by the least square method.Sufian F., Kamarudin F. & Nassir A.M., (2016), used bootstrap DEA and bootstrap regression. Wanke P., Maredza A. & Gupta R., (2017), used Network DEA and robust regression approaches such as Tobit, Simplex, and Beta.

It has been observed that sometimes in the DEA study, the variables are highly correlated and regression analysis is very sensitive to multicollinearity in data becausein such a case, regression analysis can incorrectly identify the important variables as insignificant. To overcome this problem,Ridge regression is a preferred technique than other forms of regression as it is less sensitive to the correlated predictors and minimizes the effect of multicollinearity by introducing some bias into the regression equation in order to reduce the variance of the estimator coefficients. Ridge regression uses least square objective with an added penalty and employs a trade-off between the bias and the variance in predictors.This helps in providing more accurate interpretation of each predictors' role in the model.

This paper attempts to estimate efficiency using CCR model of DEA, in a multiperiod environment, in combination with Ridge Regression using R software, with the objective of identifying determinants of efficiency, using database of Indian public sector banks (PSBs) from the year 1998 till 2013.Present study, instead of regressing DEA efficiency scores, is based on regressing output variables on input variables.

From the results of DEA, it was found that during the total time period under study, only 51 percent observations for PSBs result in efficiency. Also, the main source of inefficiency was found to be over usage of inputs 'borrowings' and 'deposits'. Results of ridge regression indicate that the input variable 'Wage Bills' is the most important determinant of both the output variables and thus the overall main determinant of efficiency.

The objective of this paper is to deal with multicollinearity of data, in a multiperiod environment so that the efficiency analysis using DEA followed by regression analysis can provide more accurate interpretations and predictions.

This paper is set out as follows. The next section describes the methodology used in this paper along with description of data and the techniques used in this study. Third section explains the empirical findings of the paper and the last section briefly lists the conclusions of this study.

II. METHODOLOGY

This section explains the research methodology been followed in this study. Subsection 2.1 provides the details of the data used in the study along with its sources. 2.2 discusses the CCR model of DEA and 2.3 explains the theoretical and mathematical process ofRidge regression followed by its use to explore dependency of individual output variables on input variables in the data envelopment analysis.

2.1 Data Base:The present studyuses the data related to 25 PSBs operating in India, for the efficiency analysis of these banks using DEA and exploring inter-dependence of variables using Ridge regression, for the period starting from the year1998 till 2013.

For the analysis, a non-parametric, input-oriented CCR model of DEA has been used, with constant returns to scale, taking four input variables and two output variables. Inputs are taken as Owned funds, Deposits, Borrowings and Wage bills. Whereas, outputs have been taken as, Spread and Other Income.

The data separately for each year, from 1997-98 to the year 2012-13, corresponding to each PSB under study, related to selected input-output variables has been obtained from the Statistical tables relating to banks in India, published by the Reserve Bank of India. Descriptive statistics of the data used in present study is given in table 1.

Table 1: Descriptive Statistics of Input & Output Variables

	N	Minimum	Maximum	Mean	Std. Deviation
Owned Funds	400	2177.0	988837.0	57197.83	98725.56
Deposits	400	47686.0	12027396.00	852906.29	1292143.55
Borrowings	400	2.0	1691827.0	51514.33	147160.26
Wage Bills	400	1289.0	183809.0	11991.01	19325.68
Spread	400	0.0	443313.0	26146.32	44157.92
Other Income	400	522.0	160348.0	10831.73	18991.88
Valid N	400				

(All variables are measured in Million Indian Rupees.)

2.2 CCR model of DEA: The CCR model of DEA, developed by **Charnes, Cooper and Rhodes (1978)**, has an input orientation and assumes constant returns to scale. It measures and compares the efficiency of decision making units (DMUs) with similar inputs and outputs. For mathematical formulation of the model, consider 'n' DMUs, each with 'm' inputs and 's' outputs, where j^{th}DMU, DMU_j , (j=1,2,….., n) uses input vector $X_j = (x_{1j}, x_{2j}, \ldots\ldots, x_{mj})$ to produce output vector

$Y_j = (y_{1j}, y_{2j}, \ldots\ldots, y_{sj})$ for $X_j \geq 0,\ Y_j \geq 0$

For input weights vector $V = (v_1, v_2, \ldots, v_m)$ and output weights vector

$U = (u_1, u_2, \ldots, u_s)$ each DMU_k has an optimization problem

$$\text{Maximize } \theta = u_1 y_{1k} + u_2 y_{2k} + \cdots + u_s y_{sk}$$

$$s.t. \qquad v_1 x_{1k} + v_2 x_{2k} + \cdots + v_m x_{mk} = 1$$

$$u_1 y_{1j} + u_2 y_{2j} + \cdots + u_s y_{sj} \leq v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj} \quad for\ all\ j = 1,2, \ldots, n.$$

$$v_1, v_2, \ldots, v_m \geq 0$$

$u_1, u_2, \ldots, u_s \geq 0 \qquad \ldots (1)$

Corresponding to $k = 1,2, \ldots, n$ (1) gives a set of 'n' optimization problems. Each problem is then solved for obtaining values of most favourable input weights $v_1, v_2, \ldots, v_m$ and output weights $u_1, u_2, \ldots, u_s$ for each corresponding DMU.

In the present study, public sector banks are considered as decision making units and CCR model of DEA is applied separately for each financial year, on the data collected for the PSBs under study, to find comparative efficiency level of each PSB in each year. Based on DEA efficiency scores, PSBs are identified to be efficient or inefficient, year wise.

2.3 Ridge Regression Analysis: Ridge regression, introduced by **Hoerl and Kennard (1970)**, introduces some bias into the regression equation to reduce the effect of multicollinearity of independent variables. In the mathematical framework, for P distinct predictor variables, the multiple linear regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_p X_p + \epsilon \qquad \ldots\ldots (2)$$

$where\ X_j\ represents\ j^{th}\ predictor\ and\ \beta_j\ gives\ the\ value\ of\ relationship\ between\ j^{th}$ $variable\ and\ dependent\ variable.$

$\beta_j\ is\ interpreted\ as\ the\ average\ effect\ of\ one\ unit\ increase\ in\ X_j\ on\ Y, keeping\ all$

$$other\ predictors\ unchanged.$$

Given estimates $\widehat{\beta_0}, \widehat{\beta_1}, \ldots\ldots, \widehat{\beta_p}$ predictions are made using the formula

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \ldots\ldots + \widehat{\beta_p}x_p \quad \ldots\ldots\ (3)$$

$\beta_0, \beta_1, \ldots\ldots, \beta_p$ *are chosen in such a way to minimize the sum of squared residuals*

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \quad \ldots\ldots (4)$$

The values $\widehat{\beta_0}, \widehat{\beta_1}, \ldots\ldots, \widehat{\beta_p}$ that minimize (3) are the multiple least squares regression coefficient estimates.

Ridge regression is very similar to least squares. The ridge regression coefficient estimates $\widehat{\beta^R}$ are the values that minimize

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}{\beta_j}^2 = RSS + \lambda\sum_{j=1}^{p}{\beta_j}^2 \ldots\ldots\ldots (5)$$

where $\lambda \geq 0$ is a tuning parameter, to be determined separately. Equation (5) trades off two criteria, minimum RSS and shrinkage penalty. Ridge regression finds coefficient estimates that fit the data well by making the RSS small, and the term $\lambda\sum_{j=1}^{p}{\beta_j}^2$, called a shrinkage penalty is small when $\beta_1, \ldots\ldots, \beta_p$ are close to zero. The tuning parameter λ serves to control the relative impact of these two terms on regression coefficient estimates. When λ=0, the penalty term has no effect and results of ridge regression will be same as the least squares estimates. Thus, the Ridge regression can be observed as an OLS regression with an additional penalty imposed.

As $\lambda \rightarrow \infty$, the impact of the shrinkage penalty grows, and the ridge regressioncoefficient estimates will approach to zero.

Unlike least squares, which generates only one set of coefficient estimates, ridge regression produces a different set of coefficient estimates $\widehat{\beta_\lambda^R}$ for each value of λ. A good value of λ is selected by using cross-validation. To perform cross validation, the dataset is randomly divided into two subsets, called the 'train set' and the 'test set'. The train set is used to calculate the coefficient estimates and these estimates are then verified on the test set.

To find the best value of λ, a grid of λ values is chosen and the cross-validation error for each of λ is computed. Then the tuning parameter value is selected, for which the cross-validation error is smallest. Then the model is re-fit using available observations and the selected value of tuning parameter.

The main advantage of Ridge regression over least squares is the bias-variance trade-off. As λ increases, the flexibility of the ridge regression fit decreases, and the shrinkage of the ridge coefficient estimate leads to a substantial reduction in the variance of predictors, at the cost of a minor increase in bias.

A ridge regressioncoefficient estimation can be also written as

$$\min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right\}\ subject\ to\ \sum_{j=1}^{p}{\beta_j}^2\ \le\ s \qquad \dots\dots(6)$$

Such that for every value of λ, there is some s, so that equations (5) and (6) will give same ridge regressioncoefficient estimates.

Present study uses Ridge regression to identify the main determinants of efficiency by regressing each output variable separately on input variables.

IIIEMPIRICAL RESULTS

This section gives the findings of efficiency analysis of PSBs under study, by using CCR model of DEA and Ridge regression. Table 2 gives the DEA efficiency scores of each bank under study, for each year, column wise.

Table 2: DEA Efficiency Scores (%)

DMUs	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Number
State	95	10	83	80	87	81	83	88	10	94	10	10	10	10	10	10	8
State	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	16
State	10	10	10	10	10	10	10	85	89	10	10	95	10	10	10	10	13
State	10	10	10	10	10	10	10	10	10	10	10	90	10	10	89	10	14
State	10	10	10	10	10	10	10	10	94	89	10	91	99	10	96	93	10
State	10	10	10	10	95	10	10	10	10	10	98	10	10	89	93	94	11
Allaha	10	10	10	95	10	10	10	10	10	10	98	10	10	94	95	88	11

Andhr	85	84	91	81	90	10	10	10	92	98	10	93	10	10	10	98	7
Bank	78	91	85	83	77	89	93	88	84	90	89	94	91	94	10	10	2
Bank	79	83	76	80	89	94	88	72	86	95	98	10	88	82	91	95	1
Bank	85	88	95	10	10	93	81	75	96	10	10	99	95	89	97	95	4
Canara	88	94	82	84	10	10	10	10	10	91	10	87	10	96	10	85	8
Centra	90	84	74	75	94	10	10	10	10	10	10	78	94	84	77	74	6
Corpor	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	16
Dena	93	87	74	69	95	10	10	84	10	10	10	10	10	10	96	93	8
Indian	59	60	62	67	82	79	96	88	89	10	10	10	10	10	10	10	7
Indian	71	82	10	10	10	10	10	10	10	10	10	98	87	86	85	88	9
Orient	95	10	10	10	10	10	10	10	10	10	99	10	10	10	10	10	14
Punjab	10	10	84	88	95	10	10	87	93	10	10	10	10	10	99	10	10
Punjab	75	81	75	88	89	10	10	10	10	10	94	91	90	78	66	79	5
Syndic	96	10	10	10	10	10	98	91	99	96	91	93	97	10	10	96	7
UCO	55	62	67	67	82	86	90	78	90	94	95	91	89	99	89	10	1
Union	88	10	58	84	10	94	87	87	10	10	10	10	10	89	10	96	8
United	71	59	52	56	76	96	10	10	97	88	69	76	10	89	92	10	4
Vijaya	68	76	87	88	95	94	10	10	10	10	89	99	99	85	78	74	4
No. of Efficie	8	12	10	10	12	16	17	14	14	16	15	11	15	12	11	11	
No. of Ineffic	17	13	15	15	13	9	8	11	11	9	10	14	10	13	14	14	
Total No. of Observations = 400; Efficient = 204 ; Inefficient = 196																	

Last column gives the number of years, out of total 16 years, for which corresponding bank is efficient. Last two rows give the number of efficient banks and number of inefficient banks year wise. To summarize, out of total 400 observations, 204 indicate efficient banks and 196 correspond to inefficient banks.

Further, Ridge Regression is employed toregress each output variable on input variablesseparately.All these calculations are done using the R software for statistical computing.For this purpose, in R, to estimate the change in coefficients depending on values of λ, a range of values of λ is chosen and regression models are built for each λ, using glmnet function from glmnet package. This function standardizes the data related to dependent as well as independent variables, before performing the regression analysis. Table 6 and Table 7 give the results of two models of Ridge regression taking dependent variables as'Spread' and 'Other Income' respectively. Independent variables in both the models are owned funds, deposits, borrowings and wage bills. To measure the accuracy of regression models, using cross-validation, respective data sets are randomly divided in two subsets, the train set (70%) and the test set (30%).

For both the models of Ridge regression, to check the data for multicollinearity, the variance inflation factors are calculated on OLS regression model for same set of variables. High values of VIF in table 3 and table 4 indicate the possibility of multicollinearity.

Table 3: Ridge Regression Model I

Dependent Variable: Spread

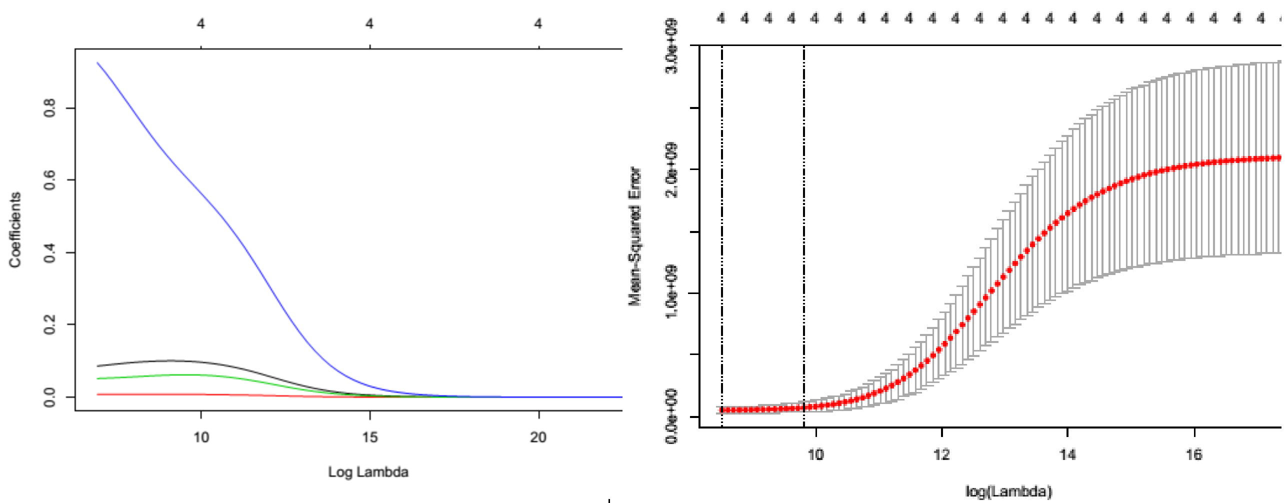
Independent Variables: Owned funds, Deposits, Borrowings and Wage Bills

Time Period: 1998-2013

Total Panel(balanced) observations: 400

	Intercept	Owned Funds	Deposits	Borrowings	Wage Bills
VIF(OLS)		73.7913	54.7044	13.5123	16.5847
$\text{Ridge}_{\lambda=4947.52}$	1637.39	0.1001	0.0081	0.0587	0.7264
For λ=4947.52	RSS_{train}=11363923385				
	RSS_{test}= 2532192842				
Accuracy = 85.67 %					

To estimate change in coefficients' value depending on λ, values of λ are taken over a sequence from three to ten, with a step size of 0.1. Figure 1(a) and 2(a) give the variation in coefficients as λ varies, for the model I and II respectively. For cross validation, Figure 1(b) and 2(b) depict the mean squared prediction error against log λ, for model I and II respectively. The results of cross validation give the best value of λ as 4947.52 for model I and 2213.92 for model II.Also, for best value of λ, Residual Sum of Squares (RSS) hasbeen calculated for train and test data sets in both models.



(a)	(b)

Fig. 1: Coefficient Estimates and Cross-Validated Estimate of Mean Squared Prediction Error, as a function of log λ, for Ridge Regression Model I

Further, comparing actual values of test data and calculated values by Ridge models, the accuracy of models is found.

Table 4: Ridge Regression Model II

Dependent Variable: Other Income

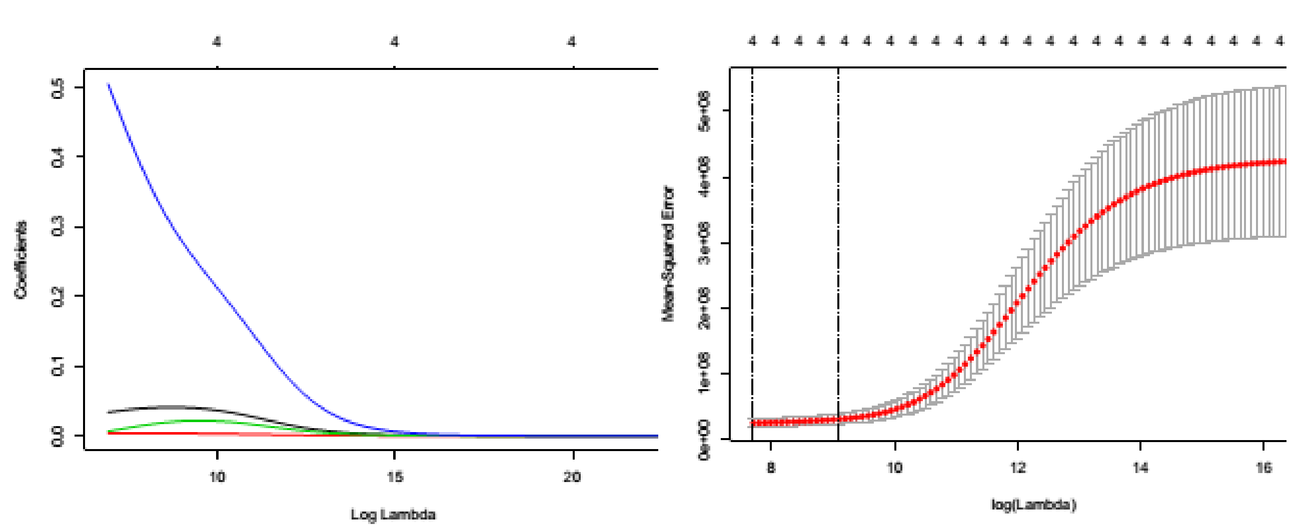
Independent Variables: Owned funds, Deposits, Borrowings and Wage Bills

Time Period: 1998-2013

Total Panel(balanced) observations: 400

	Intercept	Owned Funds	Deposits	Borrowings	Wage Bills
VIF(OLS)		63.905	55.446	14.301	16.157
Ridge$_{\lambda=2213.92}$	281.1837	0.0390	0.0035	0.0131	0.4042
For λ=2213.92	RSS_{train}= 5869795172				
	RSS_{test}= 2084103895				
Accuracy = 77.42 %					

Figure 1(a) and 2(a) show that for an increase in value ofλ, the coefficient estimates shrink towards zero. When λ is extremely large, then all the ridge coefficient estimates tend to be zero, this corresponds to the null model that contains no predictor. The upper part of these figures show the number of non-zero coefficient estimates for the corresponding value of log λ. It is noted that this number is constant for all values of logλ and is equal to the number of independent variables in the data. The dashed bars at each point show MSE_{λ} plus and minus one standard error. One of the standing dotted lines shows the location of minimum MSE and the other one shows the location of the point given by “one standard error” rule.



(a) (b)

Fig. 2: Coefficient Estimates and Cross-Validated Estimate of Mean Squared Prediction Error, as a function of log λ, for Ridge Regression Model II

From the results of Ridge regression given in table 3 and 4, it is found that the input variable ‘wage bills’ is the most prominent determinant of both the output variables ‘spread’ and ‘other income’ and has a positive relationship with both of them.

Conclusion: In the present study efficiency of 25 public sector banks operating in India has been analysed for the time period from 1998 to 2013. The comparative efficiency level of each bank under study has been evaluated using data envelopment analysis, separately for each year. Based on the efficiency levels thus found, banks are identified as efficient or inefficient for the said year. It has been found that out of total 400 observations of 25 banks for the 16 years of study, 204 result in efficiency and 196 indicate inefficiency. As the DEA results are year wise, to find overall determinants of efficiency, present study has used Ridge regression and identified the most influential input variables for each output variable. The results of Ridge regression have indicated that the input variable 'wage bills' is the most impactful input variablefor both the output variables.

References:

- Ariff M. & Can L., (2008), Cost and profit efficiency of Chinese banks: A non-parametric analysis, China Economic Review 19, 260–273.
- Assaf A.G., Barros C.P. &Matousek R., (2011), Technical efficiency in Saudi banks, Expert Systems with Applications 38, 5781–5786.
- Batir T.E., Volkman D.A. &Gungor B., (2017), Determinants of bank efficiency in Turkey: Participation banks versus conventional banks, Borsa Istanbul Review xx, 1-11.
- Charnes A., Cooper W.W. & Rhodes E., (1978), Measuring the efficiency of decision making units, European Journal of Operational Research, 2, 429-44.
- Chiu Y.H. & Chen Y.C., (2009), The analysis of Taiwanese bank efficiency: Incorporating both external environment risk and internal risk, Economic Modelling 26, 456–463.
- Chortareas G.E., Girardone C. &Ventouri A., (2012), Bank supervision, regulation, and efficiency: Evidence from the European Union, Journal of Financial Stability 8, 292–302.
- Chortareas G.E., Garza-Garcia J.G. &Girardone C., (2012), Competition, efficiency and interest rate margins in Latin American banking, International Review of Financial Analysis 24, 93–103.
- Chortareas G.E., Girardone C. &Ventouri A., (2013), Financial freedom and bank efficiency: Evidence from the European Union, Journal of Banking & Finance 37, 1223–1231.
- Das A. & Ghosh S., (2009), Financial deregulation and profit efficiency: A nonparametric analysis of Indian banks, Journal of Economics and Business 61, 509–528.
- Hoerl, A. E., & Kennard, R. W., (1970), Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 12, 55–67.
- James G., Witten D., Hastie T., Tibshirani R., (2013), An Introduction to Statistical Learning with Applications in R, Springer.

- Kamarudin F., Sufian F. & Nassir A.M., (2016), Global financial crisis, ownership and bank profit efficiency in the Bangladesh's state owned and private commercial banks, Contaduría y Administración 61, 705–745.
- Kaur R. & Aggarwal M., (2017), Complementing Data Envelopment Analysis with Regression: A New Approach,Shanlax International Journal of Management,VOLUME 5, SPECIAL ISSUE 1, vol 5, 25-39.
- Melkumova L.E. &Shatskikh S.Y., (2017), Comparing Ridge and LASSO estimators for data analysis, Procedia Engineering 201, 746-755.
- Sufian F. (2009), Determinants of bank efficiency during unstable macroeconomic environment: Empirical evidence from Malaysia, Research in International Business and Finance 23, 54–77.
- Sufian F., (2011a), An exploration into the home field, global advantage and liability of unfamiliarness hypotheses in multinational banking, IIMB Management Review 23, 163-176.
- Sufian F. (2011b), Banks total factor productivity change in a developing economy: Does ownership and origins matter? Journal of Asian Economics 22, 84–98.
- Sufian F. & Habibullah M.S., (2012), Globalization and bank efficiency nexus: Symbiosis or parasites? Review of Development Finance 2, 139–155.
- Sufian F., Kamarudin F. & Nassir A.M., (2016), Determinants of efficiency in the Malaysian banking sector: Does bank origins matter? Intellectual Economics10, 38–54.
- Wanke P., Barros C.P. &Emrouznejad A., (2016), Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks, European Journal of Operational Research 249, 378–389.
- Wanke P., Maredza A. & Gupta R., (2017), Merger and acquisitions in South African banking: A network DEA model, Research in International Business and Finance 41, 362–376.
- Yan X., Su X.G., (2009), Regression Analysis: Theory and Computing, World Scientific Publishing Co. Pte. Ltd.
- ……………………………………………………………………………………