ON DIMENSION REDUCTIONS IN DATA ENVELOPMENT ANALYSIS WITH AN ILLUSTRATIVE APPLICATION TO TURKISH BANKING SECTOR PERFORMANCE

Gizay DAVER1  Güray KÜÇÜKKOCAOĞLU2

Abstract

In this study we tested the Turkish Banking Sector Performance development and the usage of various methodologies. For this purpose we used several data envelopment analysis (DEA) performance measurement systematics from 2002.12 to 2017.06. Proposed DEA models include dimension reduction techniques. The first one is principal component analysis which is abbreviated as PCA and the second one is correlation based DEA model construction. We investigated performance change over time and found that banks in Turkish Banking Sector should take steps now for maintaining their improved performances. Policies implemented after the 2001 crisis had positive effects during the 2008 crisis period in the Turkish banking sector, but gains should be sustained. We also found that our main model which uses four input and four output factors yields similar results with the investigated dimension reduction technique approaches.

Keywords: Performance, DEA, PCA, Turkey, Financial Crisis
JEL Classification: C67, G01, G17, G18, G21 and G28

AÇIKLAYICI BİR UYGULAMA İLE TÜRK BANKACILIK SEKTÖRÜ PERFORMANSINA VERİ ZARFLAMA ANALİZİ BOYUT AZALTIMLARI

Öz


Anahtar Kelimeler: Performans, DEA, PCA, Türkiye, Finansal Kriz
JEL Siniflandırması: C67, G01, G17, G18, G21 ve G28

1 Dr.Öğr. Dyesi, Zonguldak Bülent Ecevit Üniversitesi UBYO, Bankacılık ve Finans Bölümü, E-mail: gizaydaver@gmail.com. ORCID: 0000-0001-5427-0741
2 Prof.Dr., Başkent Üniversitesi, Başkent Üniversitesi İİBF, İşletme Bölümü, E-mail: gurayk@baskent.edu.tr. ORCID: 0000-0001-6170-3269

DOI: 10.18092/ulikidince.491820
Makalenin Geliş Tarihi (Recieved Date): 03-12-2018
Yayına Kabul Tarihi (Acceptance Date): 01-05-2019
1. Introduction

Performance measurement with Data Envelopment Analysis (DEA) is becoming more accepted over time, especially in the financial services industry. There are several shortcomings of DEA despite the strengths of the methodology. When there is not enough number of Decision Making Unit (DMU)’s, factors associated with performance measurement cannot be used as required. We aimed to test explanatory inter consistency of factor reduction techniques. This objective means: If the reduced factor models capture the changes in time as the full factor model does, then detailed observations can be done by clustering the DMU’s in to smaller and more meaningful parts such as public commercial banks, private commercial banks, foreign commercial banks established in Turkey, etc. rather than a wider topic as commercial banks. With this purpose we updated our data set and used our main model stated in (Daver and Küçükkocaoğlu, 2016) which uses four input and four output factors. In this DEA study, Principal Component Analysis (PCA) methodology is implemented along with the correlation based factor reduction techniques. Particularly for deposit banks, the input factors of Deposits and Fixed Assets are found as explanatory, time saving and reasonable, as compared to any combinations of Labor, Deposits, Fixed Assets, and Bank Capital. By using PCA-DEA, at least 15 DMUs are required for capturing performance changes in time; whereas by using DEA scenarios, at least 18 DMUs are required. If all factors are used at least 24 DMU’s are required for the analyses. Factor reduction techniques are found to be meaningful and useful. This study found that PCA methodology followed by DEA application, which is known as PCA-DEA, can also be used for observing performance changes in time as well as the DEA application for the deposit banks in Turkey.

The rest of this study is organized as literature review in the second part, the methodology and models in the third part. Findings are presented in the fourth part. Summary and conclusions are drawn in the last part.

2. Literature Review

Performance is relatively new issue under the economics and finance area, and there is not a generally accepted industry standard for the performance evaluation as well. In order to capture researchers’ goals in the issue, DEA has become the most common methodology in operational research and artificial intelligence techniques (Duygun Fethi and Pasiouras, 2010). Despite of this, DEA is still young for banking and finance studies as can be seen from the citation based literature survey identifying key papers in the area (Liu, Lu, Lu, and Lin, 2013). Factor or variable selection methodology, discriminatory power of DEA, dimension reduction or factor reduction are some of the key concepts when the implementation of DEA is considered (Adler and Yazhemsks, 2010; Nataraja and Johnson, 2011; Toloo and Babaeey, 2015). In the banking and finance area, one of the most recent studies of DEA focuses on the factor selection for the retail banking industry from the output side (Eskelinen, 2017). Models proposed in the literature have a wide scale from classic DEA models to more sophisticated DEA models. There is no generally accepted standard for the best usage currently.

In the context of Turkey, the most significant works on financial institutions performance and bank performance can be listed as follows. For the periods between 2003 - 2008, CRS-DEA and data mining techniques are deployed to the Turkish banking sector data. For the data mining analyses, financial ratios are defined under clusters. Results showed that the total loan to total deposit ratio was found as a determinant of bank performance (Seyrek and Ata, 2010). Zaim and Ertugrul’s (1996) implementation of Data Envelopment Analysis used input factors of the number of employees, total interest expenses, amortization costs, and other expenses, and the output factors were the volume of short and long term deposits in Turkish Lira and the short and long term loans in Turkish Lira. Jackson, Fethi and Inal’s (1998) study used two input factors, which are the number of employees and non-labor operating expenses with output factors of loans and deposits for the periods between 1992-1996. Işık and Hassan (2003) used loanable funds, labor and capital as input factors and short term loans, long term loans, off balance sheet items and other earning assets as...
output factors from 1981 to 1990. BRSA used financial ratios as input and output factors for the purpose of DEA (BRSA, 2010). As it can be seen from these studies there is a mass of factors being used for DEA purposes. Needless to say, the search continues for the most reliable instrument and combination of factors. But there exists a much important issue the presence of enough DMUs for the used factors for the DEA.

To solve this issue, studies were done (Adler and Yazhemsky, 2010; Nataraja and Johnson, 2011; Toloo and Babaee, 2015). One of the models proposed was using PCA with DEA. PCA usage with DEA in the literature can be seen in different ways. The main two methods noted are for reducing factors being used in the DEA and for how to classify results gathered from DEA.

In the Romanian Banking Sector, PCA is used after DEA to classify banks into different groups according to their efficiency scores and to identify their business strategies. This study is similar to the ones done in the computer sciences field, which involves classification of iris flowers according to the chosen variables such as petal length, width etc. (Stoica, Mehdian, and Sargu, 2015). Another study on DEA is to reduce vertical dimensionality of certain data mining databases focused on the decision support systems. This study also addressed feature selection problems faced in the filter methods and wrapper methods. PCA falls under the filter methods and the usage of DEA methodology in this study is also for classification (Pendharkar and Troutt, 2011).

Importance of factor selection in the DEA efficiency evaluation process is mentioned in various studies. An empirical study involving commercial banks of China proposed a cash value added methodology and selection according to the statistical methods. In the study, authors discussed selection criteria and reviewed prior studies to put inconstancy of the input and the output factors of DEA evaluation framework in the banking industry briefly (Luo, Bi, and Liang, 2012). In the case study of Serbia, (Andrejić, Bojović, and Kilibarda, 2013) declared the problem to be the indicator selection for their DEA process. From previous studies, they listed a large number of indicators that have been used for efficiency measurement in the field of logistics. In order to increase the discriminatory power of DEA, they used PCA-DEA methodology. They used PCA to reduce input and output factors and then continued with DEA. For the evaluation of operational efficiency in the logistics of iron and ore at some ports, a PCA-DEA procedure was set up – two stage analysis beginning with PCA and after the reduction of factors, continuing with DEA. The reason stated was overcoming the mutual interferences related to the factors used in the study. The study found that a PCA-DEA integrated model estimates efficiency in a practical and accurate way (Chen et al., 2015).

Amongst (Cooper, Kingyens, and Paradi, 2014; Duygun Fethi and Pasiouras, 2010; Seyrek and Ata, 2010; Thanassoulis, 1999; Zhao, Casu, and Ferrari, 2009) either implementing DEA or in searching for the usage of DEA, some studies stated that classical DEA methodology did not give good results. In other words, sophisticated DEA models have higher discriminating power than classical DEA models in their studies (Andrejić et al., 2013; Zhou, Poh, and Ang, 2007). One of the shortcomings of classical static DEA was repeating periods and applying standard regression to capture efficiency change in time. Also listed was the capability of capturing the interactions between periods and computational disadvantage. With such a long time scoped historical data, for capturing efficiency changes between periods, then windows analysis would be better (Cook and Seiford, 2009). Advancing computer science can overcome the computational disadvantages with coding. In this study, windows analyses are not used, but models are tested for internal consistency.

In this study there are two interconnected problematic issues. The originality and the contribution of this paper can be observed by looking at the methodology to overcome these issues. First, the number of input and output factors. And the second, is the number and status changes of DMUs that are observed under a group. While trying to resolve these issues we concentrate on observing performance change difference throughout analysis period between the models that are used.
3. The Methodology and Models

In order to test the performance of banking system in Turkey, data envelopment analyses were implemented. Data set is gathered from the Banks Association of Turkey. At first, stage four models were tested according to the correlation combinations. In the second stage, for input side factor reduction, PCA is completed under SPSS software. Then according to the findings, new DEA models are proposed for each quarter. In the last stage, DEA and PCA-DEA scores are compared.

3.1. Data Envelopment Analysis and Principal Component Analysis

Quarterly input factors were labor, deposits, fixed assets and bank capital, and quarterly output factors were non-performing loans, securities, off balance sheet assets and loans (Daver and Küçükkocaoğlu, 2016). Within the output side, non-performing loans are associated with bad luck, bad management and skimping (Berger and Mester, 1997). Impacts of negative shocks, assumed as exogenous factors are named “bad luck”, and wrong construction of bank’s loan portfolios are named as “bad management”. According to the skimping hypothesis, non-performing loans are associated with mismanagement of control and loan monitoring. The rationale of these input output factor selections involves back testing for the previous study cited, not for any hypothesis testing.

Four inputs and four outputs were used in the main model. Banks use personnel, fixed assets, bank capital and deposits to create the desired and undesired output throughout their intermediation function. Securities, off balance sheet assets and loans are the ones that are favorable. The non-performing loans are the unfavorable ones.

Table 1: DEA Input Correlations for Scenario Building

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>Deposits</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td>M2</td>
</tr>
<tr>
<td>Fixed Assets</td>
<td>0.85</td>
<td>0.85</td>
<td>1</td>
<td></td>
<td>M3</td>
</tr>
<tr>
<td>Bank Capital</td>
<td>0.95</td>
<td>0.99</td>
<td>0.82</td>
<td>1</td>
<td>M4</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

There are status and/or name changes through the period, 2014 name and status are used for the purpose of the analyses. By only looking at the main groups, the Turkish banking sector is clustered under three DMU groups. These clusters are the whole banking system, deposit banks, development and investment banks. From this point the Turkish banking sector is not eligible to develop any of the DEA models used in this study. One of the three clusters of the Turkish banking sector, which is the whole banking system DMU contains both of the other two DMU’s. When clusters are detailed as state owned, privately owned and foreign, this result would not even be enough for the analyses. Eight input/output variables require a minimum of sixteen DMU’s to result in reasonable efficiency scores, whereby the data in this study has more than triple the variables. The number of DMU’s used in this study is twenty nine. Another issue is the chosen input and output variables. Development and investment banks are automatically excluded from the analyses. From an input oriented approach, three more models are developed by looking at the 2014/4th quarters input factor correlations. This has resulted in four models coded as M1, M2, M3 and M4. In order to give an illustrative example, model number one(M1) which uses four input and four output factors and model number two(M2) which uses two input and four output factors are
compared. In M1, input factors are personnel(labor), fixed assets, bank capital and deposits. Output factors are securities, off balance sheet assets, loans and the non-performing loans. In M2, input factors are personnel(labor) and fixed assets. Output factors are securities, off balance sheet assets, loans and the non-performing loans. Correlation matrix and models variables are on Table 1. Four output factors are used in all models. Transformations on data are done considering, translation invariance, scale invariance, negative data treatment and zero data treatment. These methods are explained and argued in previous studies (Fried, Lovell, and Schmidt, 2008; Ray, 2004). All of these models are solely dependent on the DEA procedure.

Efficiency measurement system software uses Andersen and Petersen’s procedure in super efficiency procedure. And big yielding results mean the DMU remains efficient under a large arbitrary increased input.

Taking these models one step further, we used PCA on the input factors of M1 with the help of SPSS software. By itself, DEA takes in to consideration only one point in time, so the interpreted 59 PCA-DEA models were set up for each quarterly DEA data from 2002.12 to 2017.06. All the input side factors yielded one principal component with different coefficients. According to PCA results, one input and four output DEA models are constructed.

4. The Findings

Results of the DEA are presented in Table 2. According to all models, we found an increased number of efficient DMUs in the 2002.12 to 2004.12 period. In 2005, inefficient DMUs increase from Models 1 and 2, causing the number of efficient DMUs to return to the 2002.12 numbers. This temporary situation ends after 2005.12, showing a recovery of efficient number of DMUs and continuing until 2011. The banking sector restructuring program and the economic recovery program helped Turkey through the 2008 crisis. In all models, fluctuations begin with the year 2011 and continue to the end of analysis period.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>PCA-DEA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big</td>
<td>Inefficient</td>
<td>Efficient</td>
<td>Big</td>
<td>Inefficient</td>
</tr>
<tr>
<td>2002/12</td>
<td>10</td>
<td>7</td>
<td>12</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>2003/03</td>
<td>10</td>
<td>5</td>
<td>14</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>2003/06</td>
<td>7</td>
<td>4</td>
<td>18</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2003/09</td>
<td>11</td>
<td>3</td>
<td>15</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>2003/12</td>
<td>6</td>
<td>6</td>
<td>17</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>2004/03</td>
<td>6</td>
<td>6</td>
<td>17</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>2004/06</td>
<td>6</td>
<td>5</td>
<td>18</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>2004/09</td>
<td>6</td>
<td>8</td>
<td>15</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>2004/12</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2005/03</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>2005/06</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>2005/09</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>2005/12</td>
<td>6</td>
<td>11</td>
<td>12</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>2006/03</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>M1 Big inefficient</td>
<td>M1 Efficient</td>
<td>M2 Big inefficient</td>
<td>M2 Efficient</td>
<td>M3 Big inefficient</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>2006/06</td>
<td>6 8 15 6 17 6 6 10 13</td>
<td>6 6 11 12</td>
<td>5 10 14</td>
<td>2006/09</td>
<td>6 6 17 6 15 8 6 10 13</td>
</tr>
</tbody>
</table>
According to Model 1, after 2010 the number of inefficient firms doubles by 2012.03 and slightly declines towards 2017 with some fluctuations. According to Model 2, especially after 2012, the number of inefficient firms increases, whereas number of efficient firms drops to about half the values seen in 2010 and 2011. Model 3 shows the numbers of efficient and inefficient firms are about the same after 2011.12 till 2015. Model 4 draws a moderate scenario; numbers of inefficient firms are higher than the 2008 crisis period after 2010.12. The worst scenario is found in Model 2 after PCA-DEA and averaging the period.

### Table 3: 2002.12 – 2017.06 PCA-DEA and DEA Model Correlation Results

<table>
<thead>
<tr>
<th></th>
<th>PCA-DEA vs M1</th>
<th>PCA-DEA vs M2</th>
<th>PCA-DEA vs M3</th>
<th>PCA-DEA vs M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>Inefficient</td>
<td>0.57</td>
<td>0.62</td>
<td>0.63</td>
<td>0.44</td>
</tr>
<tr>
<td>Efficient</td>
<td>0.58</td>
<td>0.63</td>
<td>0.74</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

In the Table 3 correlation results are given. PCA-DEA —applying principal component analysis at first, and then running DEA by using the principal component as the input factor—are presented versus DEA models M1, M2, M3 and M4. Strong correlation results of M1, M2 and M3 models of DEA with PCA-DEA showed that PCA-DEA can explain the performance change over time by looking at the number of efficient or inefficient DMUs. Results emphasizes that PCA-DEA does not capture the performance changes of M4 over time in a strong state.

### Summary and Conclusions

In the analysis period from the selected deposit banks, from DEA model 1 and DEA model 4 (with negative signs), nearly no change in the efficiency trend is observed. From model 2, slightly more efficiency decline is observed. In model 3, a result contrasting to other models is observed, stating the efficiency increase with a positive sign. Though the sign is positive, this model indicates a nearly no change in the efficiency trend in selected deposit banks. One of the most important findings is that DEA results of proposed models using different input combinations, but the same output set, yield different results. When all inputs, such as labor, deposits, fixed assets and bank capital are used, it is found that nearly no efficiency change in the analysis period is observed. Model 2 and model 4 states efficiency decline in time similarly, however a sharper efficiency decline in fixed asset and bank capital usage is observed than the labor and fixed asset usage.
Another conclusion from the correlation results of PCA-DEA with DEA models is that researchers may have a reasonable justification to focus on the input factors of Deposits and Fixed Assets. We also conclude that these two factors have more explanatory power in the comparative analyses.

With PCA-DEA, the total factors are reduced to five. This dimension reduction allows us to use fifteen DMUs (actual banks or clusters). By using a scenario building approach, we need eighteen DMUs in models M2 to M4. We also require six more banks in M1. The strong correlation between methods shows that we can reduce factors by using PCA approach. If we want to use least factors solely, Deposits and Fixed Assets would be better than any combinations of Labor, Deposits, Fixed Assets and Bank Capital. Dimension reduction can be applied in Turkish banking sector, but because of the classification of Banking Regulation and Supervision Agency’s or The Banks Association of Turkey, the adequate number of DMUs (clusters) could not be reached, even after factor reduction.

Today financial data is large compared to other areas of science. And calculations become a critical factor for researchers. This study showed that factor reduction techniques can be implemented. The prerequisites of the proposed methodologies and/or models are the constraints and computer assisted calculations are more essential for the data used in financial studies. Future studies in opening codes and showing calculation steps are recommended along with related transparency planning.

References


