10
Online Banking Efficiency and Risk Evaluation with Principal Component Analysis

Introduction

Online banking has always been important from different stakeholder perspectives.[1] Improving the efficiency of Internet banking is now considered to be important. Banks hope that internet banking will help them maintain profitable growth by enabling them to automate work, reduce costs, and retain customers simultaneously.[2] Internet banking may help reduce expenditure on 'bricks and mortar,' and reduce capital expenditures.[3] Internet banking can give customers 24-hour access, and provide convenience for customers. Cost-effective use of the internet can attract many users to online banking services, but there has been little research examining the superiority of banks providing online banking services over those that do not. The role of online banking in leading to better decisions and creating more profit needs study.

The Economist has reported that online banking costs were approximately \$0.01 per transaction in 2000, accounting for only 1% of all banking transactions.[4] Online banking costs are increasing rapidly. Jupiter[5] reported that there were an estimated 30 million families in the US using internet banking in 2004, and forecast that the numbers would reach 56 million in 2008. The growth of internet banking is due to economic globalization and the maturation of computer technology. Economic globalization encourages banks to serve and exploit international markets. Opening up a new virtual bank branch requires high expenditure and faces limitations, but online banking websites can provide services to customers throughout the world as long as customers are able to surf the internet. The reduction of computer prices makes their purchase accessible to most people. Meanwhile, perceived improvement in computer security makes customers and banks sufficiently comfortable to use them for their business. The consequent rapid increase of online banking transactions creates an urgent need to examine the efficiency of online banking.

Furthermore, efficiency evaluation is a source of ideas for bank managers, and motivates them to improve the quality of online banking services.

Serrano-Cinca et al. (2005)[6] suggested that financial information alone might not always be sufficient to judge an online business. Therefore, constructing performance evaluation methods that can combine a number of possible inputs and outputs is attractive. Banks can evaluate their online banking performance using online performance measurement considering an efficient frontier of tradeoffs across selected criteria. Previous literature on online banking performance evaluation includes the following approaches: linear regression (e.g. the logit model in Furst et al.), DEA (e.g. Sherman and Gold, 1985[7]; Soteriou and Zenios, 1999[8]), free disposal hull (e.g., Tulkens, 1993[9]), the stochastic frontier approach (also called econometric frontier approach, e.g. Berger and Humphrey, 1992[10]), the thick frontier approach (e.g. Berger and Humphrey, 1997[11]), the distribution free approach (e.g., Berger et al., 1993[12]), and others. The main differences between these approaches lie in how much restriction is imposed on the specification of the best practice frontier and the assumptions of random errors and inefficiency.[13] DEA is recognized as useful for performance analysis in banking industry because of its advantages in allowing for dynamic efficiency without requirements for prior assumptions on the specification of the best practice frontier. This translates into practical advantages for DEA over other methods, in that an explicit specification of a mathematical form for the production function is not needed, and DEA can provide an integrated efficiency score to decide the level of efficiency of a specific bank. Conversely, there can be computational difficulty if there are too many input and output variables.

Principal component analysis (PCA) is a flexible approach that can be used to reduce the number of variables and to classify independent variables. PCA can thus provide powerful support to DEA by providing a means to reduce the number of input variables and integrating output variables. PCA has been applied in many scientific disciplines, including statistics, economics, finance, biology, physics, chemistry, and so on. In internet banking, Eriksson et al. (2008)[14] used PCA to identify determinants of customer satisfaction. PCA has also been used to study the acceptance of internet banking by customers, with a primary conclusion suggesting that banks place effort on making internet banking user-friendly.

Data and variables

To analyze efficiency, we looked into a few banks chosen from the UK and the US. Data was gathered from 2007 annual reports (see the reference list for such reports) of these banks divided into the input variables and the output variables. Table 10.1 gives input and output variables:

Table 10.2 presents online banking data for ten large banks, six from the US and four from the UK: Bank of America,[15] Citigroup,[16] HSBC,[17] Barclays,[18]

Table 10.1 Online banking DEA variables

INPUT variables	Designator	OUTPUT variables	Designator
Total deposits	A	Total revenue	1
Operating cost	B	Daily visits	2
Employees	C		
Equipment	D		

Table 10.2 Online banking data

Bank	Total deposits	Operating cost	Employees	Equipment	Revenue	Daily reach
Bank of America	805177	89881	210000	9404	68068	4524000
Citigroup	826230	61488	387000	8191	81698	623000
HSBC	278693	39042	330000	6054	87601	136000
Barclays	657058	26398	135000	5992	40410	1083000
Chase	740728	110560	180667	3779	71400	2030000
Wells Fargo	344460	22824	159800	1294	39390	1051000
Lloyds	393092	11134	70000	4014	58180	623000
Royal Bank of Scotland	595908	28106	226400	3524	108934	134000
SunTrust	119877	5234	32323	1739	8251	129000
Wachovia	449120	9465	29940	6605	17653	551000

Chase,[19] Wells Fargo,[20] Lloyds,[21] Royal Bank of Scotland,[22] SunTrust,[23] and Wachovia.[24]

Results and analysis

We first conducted a plain DEA analysis and then used PCA scenario analysis and PCA-DEA modeling.

Table 10.3 presents the scores produced from normal DEA models based on all 45 combinations of variables in the data. A score of 1.000 (in bold) indicates efficiency. It can be seen in the ABDCD12 column that seven of the ten banks are found efficient. The problem is that using more variables obviously finds more efficient solutions. The ABCD12 model generates too many efficient DMUs/ties due to many input and output variables. But many variables and factors can affect each other in real world; for example, operating cost is usually related to the number of employees in the corporation. Therefore PCA is applied to reduce these measures when applying DEA.

To conduct PCA analysis, we use various combinations of variables to see if we can reduce the number of variables and/or detect structure in the relationships between variables. Table 10.4 presents maximum component loadings matrix in different models with different data. All models are weighted with a positive sign

Table 10.3 DEA combinations and their efficiencies

A1	B1	C1	D1	AB1	AC1	AD1	BC1	BD1	CD1	ABC1	ABD1	ACD1	BCD1	ABCD1
0.269	0.145	0.390	0.234	0.309	0.521	0.366	0.390	0.234	0.447	0.521	0.366	0.521	0.447	0.521
0.315	0.254	0.254	0.323	0.447	0.501	0.458	0.254	0.339	0.400	0.501	0.458	0.501	0.400	0.501
1.000	0.429	0.319	0.468	**1.000**	**1.000**	**1.000**	0.429	0.555	0.526	**1.000**	**1.000**	**1.000**	0.555	**1.000**
0.196	0.293	0.360	0.218	0.362	0.403	0.294	0.360	0.343	0.414	0.403	0.362	0.414	0.414	0.414
0.307	0.124	0.476	0.611	0.307	0.605	0.611	0.476	0.611	0.751	0.605	0.611	0.751	0.751	0.751
0.364	0.330	0.297	0.985	0.545	0.582	0.985	0.330	0.985	0.985	0.582	0.985	0.985	0.985	0.985
0.471	**1.000**	**1.000**	0.469	**1.000**	**1.000**	0.677	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.582	0.742	0.579	**1.000**	**1.000**	**1.000**	**1.000**	0.742	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.219	0.302	0.307	0.154	0.390	0.420	0.271	0.307	0.311	0.319	0.420	0.390	0.420	0.319	0.420
0.125	0.357	0.709	0.087	0.357	0.709	0.154	0.709	0.357	0.709	0.709	0.357	0.709	0.709	0.709
A2	**B2**	**C2**	**D2**	**AB2**	**AC2**	**AD2**	**BC2**	**BD2**	**CD2**	**ABC2**	**ABD2**	**ACD2**	**BCD2**	**ABCD2**
1.000	0.865	**1.000**	0.592	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.134	0.174	0.075	0.094	0.198	0.134	0.149	0.174	0.199	0.135	0.198	0.199	0.149	0.199	0.199
0.087	0.060	0.019	0.028	0.087	0.087	0.087	0.060	0.068	0.038	0.087	0.087	0.087	0.068	0.087
0.293	0.705	0.372	0.223	0.764	0.372	0.344	0.705	0.774	0.375	0.764	0.774	0.375	0.774	0.774
0.488	0.315	0.522	0.661	0.488	0.522	0.786	0.522	0.661	0.952	0.522	0.786	0.952	0.952	0.952
0.543	0.791	0.305	**1.000**	0.892	0.543	**1.000**	0.791	**1.000**	**1.000**	0.892	**1.000**	**1.000**	**1.000**	**1.000**
0.282	0.961	0.413	0.191	**1.000**	0.413	0.308	0.961	**1.000**	0.413	**1.000**	**1.000**	0.413	**1.000**	**1.000**
0.040	0.082	0.028	0.047	0.090	0.040	0.060	0.082	0.094	0.062	0.090	0.094	0.062	0.094	0.094
0.192	0.423	0.185	0.091	0.462	0.192	0.192	0.423	0.445	0.185	0.462	0.462	0.192	0.445	0.462
0.218	**1.000**	0.854	0.103	**1.000**	0.854	0.218	**1.000**	**1.000**	0.854	**1.000**	**1.000**	0.854	**1.000**	**1.000**
A12	**B12**	**C12**	**D12**	**AB12**	**AC12**	**AD12**	**BC12**	**BD12**	**CD12**	**ABC12**	**ABD12**	**ACD12**	**BCD12**	**ABCD12**
1.000	0.865	**1.000**	0.592	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.395	0.254	0.254	0.324	0.492	0.514	0.507	0.254	0.399	0.423	0.514	0.525	0.542	0.434	0.542
1.000	0.429	0.319	0.468	**1.000**	**1.000**	**1.000**	0.429	0.555	0.526	**1.000**	**1.000**	**1.000**	0.555	**1.000**
0.403	0.707	0.476	0.223	0.764	0.522	0.462	0.707	0.774	0.558	0.764	0.774	0.558	0.774	0.774
0.652	0.316	0.644	0.661	0.652	0.789	0.834	0.644	0.661	**1.000**	0.789	0.834	**1.000**	**1.000**	**1.000**
0.747	0.794	0.391	**1.000**	0.924	0.783	**1.000**	0.794	**1.000**	**1.000**	0.924	**1.000**	**1.000**	**1.000**	**1.000**
0.651	**1.000**	**1.000**	0.472	**1.000**	**1.000**	0.786	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.582	0.742	0.579	**1.000**	**1.000**	**1.000**	**1.000**	0.742	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
0.348	0.432	0.337	0.155	0.488	0.466	0.365	0.432	0.445	0.353	0.488	0.488	0.466	0.445	0.488
0.280	**1.000**	**1.000**	0.103	**1.000**	**1.000**	0.280	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Notes: A: Deposit; B: Operation cost; C: The number of employees; D: Equipment; 1: Revenue; 2: Web reaches.

on the first component and other variables, thus, the first component is named 'overall measure of efficiency,' and is the higher-weighted value in general. In all models, variables can be reduced so that three principal components can be used to explain more than 85% of the variation in the raw data.

In this research we chose the principal components with the accumulative contribution ratio equal to or above 90%, and applied them to reevaluate the efficiency of the online banking. For instance, in model ABCD12 we can conclude that only three principal components are used, because the accumulative contribution of three principal components is 90.2%. The calculation is easy due to the reduction of data complexity, and the ranks are reasonable.

The data extraction process is shown in Figure 10.1, where every vector represents each of the six variables. How each variable contributes to the three principal components is indicated by the direction and length of the vector.

Table 10.4 Maximum component loadings matrix in different models

Model	**PC1**	**PC2**	**PC3**
ABCD1	0.851	–0.561	0.504
ABCD2	0.895	0.750	–0.598
ABC12	0.892	0.719	–0.467
ABD12	0.906	0.804	0.640
ACD12	0.868	0.705	–0.526
BCD12	0.849	0.739	0.662
ABCD12	0.881	–0.675	0.674

Notes: A: Deposit; B: Operation cost; C: Number of employees; D: Equipment; 1: Revenue; 2: Web reaches.

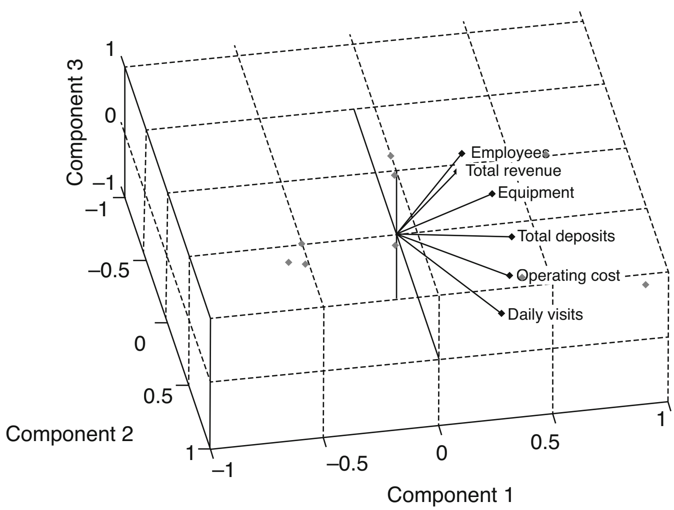

Figure 10.1 3D plot of PCA analysis (ABCD12 model)

In the three-dimensional solid area it is shown that the first principal component, represented by the Component 1 axis, has positive coefficients for all six variables. The second principal component, represented by the Component 2 axis, has negative coefficients for variables Total deposits (A), Operating cost (B), Equipment (D) and Daily visits (2), and positive coefficients for the remaining two variables. The Component 3 axis has all positive and negative coefficients for the variables. This figure indicates that these components distinguish between online banking that has high values for the three sets of variables, and low for the rest. This approach can effectively achieve dimensionality reduction without losing too much information.

PCA-DEA analysis

This section analyzes online banking using an integrated PCA-DEA model. As internet usage grew, it considerably changed the channel between banks and clients. We used both financial and non-financial variables. The main objective is to construct a framework with DEA and PCA approaches, and use it for the measurement of online banking based on data collected from annual reports and web metrics.

The basic DEA model in Table 10.2 shows us that we have seven banks that are efficient. To reduce the number of variables and obtain more accurate results, we ran a PCA-DEA analysis at 75%. This left us with three efficient banks: Bank of America, Lloyds and the Royal Bank of Scotland. Table 10.5 presents integrated PCA-DEA scores, and Table 10.6 gives variance explained by integrated PCA-DEA.

Only the Bank of America gets 100% efficiency in all models. Note that in Model ABCD1, where all inputs and outputs except for Daily Reach are taken

Table 10.5 Integrated PCA-DEA score

PCA-DEA – 75%	**ABD12**	**ABC12**	**ABCD1**	**ACD12**	**ABCD12**
Bank of America	**100%**	**100%**	**100%**	**100%**	**100%**
Citigroup	48%	43%	14%	50%	48%
HSBC	86%	80%	5%	76%	75%
Barclays	47%	56%	39%	53%	57%
Chase	65%	60%	54%	99%	77%
Wells Fargo	100%	75%	67%	100%	98%
Lloyds	**87%**	**100%**	**38%**	**97%**	**100%**
Royal Bank of Scotland	**100%**	**94%**	**5%**	**100%**	**100%**
SunTrust	36%	45%	20%	39%	41%
Wachovia	29%	50%	28%	32%	38%
Min	29%	43%	5%	32%	38%
Max	100%	100%	100%	100%	100%
Mean	70%	70%	37%	75%	73%
Standard Deviation	0.2808	0.2239	0.3005	0.2805	0.2579

into account, only one bank gets a score of 100% and the average efficiency is only 37%. In the other models where the daily reach is taken into account the average ranges from 70 to 75%.

To analyze all efficiency scores in Table 10.2, we ran PCA on all 45 sets of scores for all banks. Figure 10.2 gives a plot of principal component loadings in different DEA models, where different scores can be understood from a different perspective. The plots of principal component loadings are converted from the matrix of component loadings and show a set of directional vectors. The computed component loadings result in meaningful naming for both horizontal (the first component) and vertical axis (the second component). The horizontal axis in Figure 10.2 is from west to east, representing the 'overall

Table 10.6 Variance explained in integrated PCA-DEA

Variance	ABD12	ABC12	ABCD1	ACD12	ABCD12
A	65.53	73.66	71.45	65.53	68.21
B	15.05	19.39	19.65	15.05	
C	14.39	0.00	8.89	14.39	20.02
D	5.04	6.95	0.00	5.04	11.77
1	53.85	53.85	53.85	100.00	53.85
2	46.15	46.15	46.15	0.00	46.15

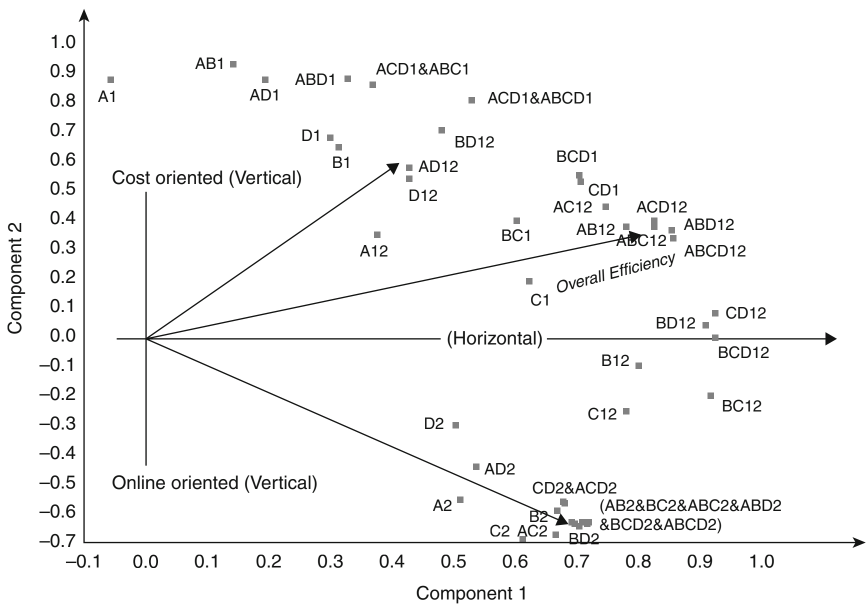

Figure 10.2 Plots of principal component loadings in different DEA models

Table 10.7 Multivariate linear regression analysis – DV bank revenue

	Standardized coefficients	T	Sig.	Correlation coefficients
Constant	NA	0.710	0.517	
A TotalDeposits	0.319	0.632	0.561	0.487
B Cost	−0.011	−0.017	0.987	0.473
C Employees	0.749	1.698	0.165	0.783
D Equipment	−0.246	−0.553	0.610	0.275
2 Daily Reach	−0.030	−0.056	0.958	0.077

measure of efficiency'; the more efficient ones overall will be located in the right direction. From the origin to north and south are the 'cost oriented' and 'online oriented' models respectively. Interestingly enough, such a finding is consistent with existing work based on data from other nations.

Risk factors

This section seeks to detect the key variables that contribute the most to bank revenue. We ran both multivariate linear regression and correlation analysis and present computed values in Table 10.7. It can be seen that the Employees variable has the largest effect on revenue with a regression coefficient value of 0.749 and correlation coefficient value of 0.783. This means that allocation of Employees will affect profit or loss the most, holding the other variables constant. That is to say, misutilization of Employees will bring huge potential risks, which should be considered by banks' managers when they decide to enter the online banking market. This is because the risks have become important factors affecting the survival of enterprises. Meanwhile, the Basel Committee on Banking Supervision requires that every bank must have an efficient regulation and risk management system where enterprise risk managers should take an important role on the board of directors.[25] Three other variables, that is, total deposits, cost, and equipment, have less effect on revenue, and daily reach has no effect on revenue at all. It may be easy to understand that the people who click the online banks' websites do not necessarily conduct transactions over the websites. Therefore, future research should examine whether the number of transactions will significantly affect revenue.

Conclusions

Various financial and non-financial variables have been used in this study to analyze the online banking service of some giant banks. PCA is employed to identify the variables contributing the most information content for DEA models. The results enable identification of the most efficient banks in terms of the particular variables selected through DEA. This information is then further analyzed in terms of multivariate linear regression, which enables significance and correlation to be seen across variables.

The combination of models applied to banking risk management issues (in this case, online banking) can provide useful tools to benchmark banking operations and to identify opportunities for improvement in those operations.