Stochastic Frontiers. Case Study – Japanese Banking System

Ionut-Cristian Ivan

Institute for Doctoral Studies, Academy of Economic Studies Bucharest, Romania

Abstract

The paper aims to calculate the average estimate of efficiency of Japanese banking system over a period of time, being given certain inputs and outputs that describe the main activities of analyzed banks. Stochastic Frontier Analysis have been used to calculate the scores, also allowing a comparison with other comparable techniques (Data Envelopment Analysis, COLS). I expect to come to average values of efficiency similar to studies conducted before this research.

Keywords: Efficiency; Stochastic Frontier Analysis; DEA

Introduction

Japanese banking environment is a system easily differentiable from other banking systems, due to its particularities, such as: presence of keiretsu banks (banks that operates only for a large conglomerate of firms), the degree of openness towards outside markets, the classification of banks in inter-regional and intra-regional banks etc.

The following research brings new information on Japanese banking system, succeeding in summarizing the associated efficiency level. Also, to calculate de efficiency estimates for each DMU (decision making unit), I used parametric methodology (stochastic frontier analysis).

If nonparametric techniques are much closer to the idea of operational research, parametric methodology lies closer to econometrics. Obtained estimators should check the assumptions of econometric estimators.

*Corresponding author.
E-mail address: cristian.ivan@csie.ase.ro (I. C. Ivan)
The article is structured as follows. The second section presents in a simple and structured way the main notions that were used to calculate the estimated parametric measures of efficiency. The third section is dedicated to the application of the aforementioned methodology. Last section concludes my research.

1. Methodology

Unlike nonparametric methodology, parametric methodology is defined by an a priori known structure of the feasible production set and generation of data, except for some parameters that form the functional form of the relationship between inputs and outputs, also known in advance. Another major difference is that in case of parametric techniques the shifts from the frontier are not due only to inefficiency, but also to white noise.

Initially, after listing the specifications associated with parametric techniques, the use of regression to estimate the frontier seems convenient. The main problem of this method is that there will be firms that will be situated outside the frontier, which violates the assumption that the frontier represents the maximum coverage of the production set, given the input and output vectors. The COLS method implies that deviations from the regression line are given by the inefficiency, while the white noise is not considered. COLS method involves the estimation of the functional parameters associated with the production process, through a regression line and then, the correction of the constant term so as to ensure that all companies are under the estimated frontier.

SFA methodology, however, combines in the model, both the error term and inefficiency. We find the following functional form in Bogetoft and Otto (2010):

\[ y_k = f(x_k; \beta) + \varepsilon = v_k - u_k \tag{1} \]

where \( v \) is the error term that follows a normal distribution with zero mean and \( \sigma_v^2 \) standard deviation and \( u \) is the inefficiency. The term \( u \) is positive and a null value indicates perfect efficiency (the analyzed firm lies on the frontier).

The main problem for this formulation is the decomposition of \( \varepsilon \) in inefficiency and white noise.

\[ \sigma^2 = \sigma_v^2 + \sigma_u^2 \tag{2} \]

\[ \lambda = \frac{\sigma_u^2}{\sqrt{\sigma_v^2}} \tag{3} \]

SFA methodology uses these terms to obtain estimated measures of inefficiency, through the maximization of the likelihood function.

In Bogetoft and Otto (2010) the following formula is presented to calculate the estimator of the efficiency measure:

\[ TE_k(x_k, y_k) = \exp(-\lambda^k) \tag{4} \]

2. Empirical Results

In order to study the efficiency of the Japanese banking system I decided to apply parametric analysis techniques. Considering the profit approach, defined by Fethi and Pasious (2009), I have chosen for the analysis the following inputs and outputs, using records from the income statement of each bank.

Based on these considerations, I have chosen as input variables - provisioning expenses, interest paid, fees and commissioning expenses and as output variables - net income, received interest and fee and commissioning revenue.

These variables were chosen as they describe the main activities of banks (making deposits - interest expenses, granting loans - interest paid and provisions made, taxes and fees associated to various banking services). Also, the provisioning expenses succeed in the incorporation of banks acumen to deal with the risk of bad loans. Net income summarizes banks' performance for a given period of time.

The data is covering 2010-2012 accounting years and 99 local banks, including the three mega-banks of Japan, hereinafter referred to as shikin banks. The banks are distinguishable by their degree of openness, whereas we included in the analysis both banks operating within a single prefectures and banks operating abroad. In general, inter-regional banks have higher values of indicators than intra-regional banks; these banks are likely to adversely affect the efficiency scores or behave as super-efficient banks. Also, intra-regional banks have associated costs of provisioning much lower than inter-regional banks, largely because rather low volume of granted loans. At the same time, the main negative net income cases are found in intra-regional banks.

For a better understanding of the links between the selected variables, I propose a brief description of each and every variable.
• Interest paid - is the cost accepted by commercial banks in order to form deposits for households and firms and it depends largely on the passive interest level controlled by the central bank and interbank deposits interest.

It also includes the cost of financial instruments interests such as securities. A full analysis of this component requires historical tracking of interest describing domestic and international market and also of interest rates adjusted by the National Bank of Japan.

We observe that, on average, interest costs decreased significantly in 2010-2012 period, with more than 5 billion yen. This may be due to losses from bad financial instruments or to decreased consumer confidence in the banking system (decide not to make savings deposits). Another explanation for the decrease in interest costs is that individuals and companies choose to increase consumption rather than savings, given the unstable economy conditions.

• Expenses with fees and commissions

Basically, this variable is composed of commissions and fees charged in inter-banking relationship. They are also charged as a result of participation in investing in financial instruments.

Analyzing the given data set, we see a slight increase in the variable cost with fees and commissions. Much of the increase is due to shikin banks, which have major variable increases during 2010-2012. The remaining banks either keep a constant level or they lower the expenses with fees and commissions. The most likely reason for the behavior of small and medium banks is uncertainty in investing in derivatives market, given the latest financial crisis. The actors become more risk adverse and they choose to invest in assets with low risk, even if their profitability is not always very high.

• Provisioning expenses

Provisioning expenses is an important indicator of the financial risk that a bank can assume. The provision represents a non-cash amount of money that a bank reserves in the profit and loss account to cover future bad loans. Calculation of reserves helps to establish the level of charges subsequent to the date of provision establishment and also helps to maintain the solvency of the bank if bad loans forecasts come true.

The evolution of the amounts withheld as provisions describes very well the behavior of banks towards the risk present in the banking environment. A high level of bank reserves describes a risk adverse bank, while a low level describes a bank with risk preference.

In 2010-2012, it can be observed a strong decrease in the average level of provisions. The trend is maintained individually for each bank. The decrease can be attributed either to improvement of the financial situation (banks considered a low risk of bad loans) or a decrease in loans to individuals or companies. I tend to choose the first embodiment, since during the last period there was an increase in investment in Japan, investments financed from loans taken from local banks.

• Interest revenues

This variable includes money obtained from borrowing by individuals and companies. Also part of interest income is derived from inter-banking loans and derivatives investment. Since investments based on loans granted by banks increased, according to recent reports of the Bank of Japan, interest income should increase as well.

Despite our expectations, the net interest income decreased on average by 10 billion yen. Although companies began to borrow for investments, their growth fails to cover the decrease in interest income obtained from the derivatives market.

• Fees and commissions revenues

This variable is formed following the receipt of commissions from regular customers in exchange for various financial services provided by the bank. Also it contains taxes withheld from working with other banks in the market.

It can be seen a slight increase in the average level of income from fees and commissions. Given the decrease in the level of loans, lowering fees and charges makes sense.

• Net Income

Net income summarizes the overall result of a bank during an analyzed period. It is obtained by subtracting depreciation, interest expenses and fees and expenses of the gross profit.

We observe a strong increase in average net income of almost 100% compared to 2010. Together with the decrease in provisions, interest expenditure and increased tax revenues and fees, it seems to exist upstream other factors also that lead to this increase (example, changes in the economic environment - low levels at the beginning of the economic crisis, followed by increases with the application of various countermeasures, by the Central Bank).
As a generalization on all analyzed variables, there is a large difference between the following values: minimum, average and maximum. This gap arises from the inclusion in the analysis of shikin banks, which have extremely high values associated to variables. This gap supports the idea of outlier existence or super-efficient banks, according to the methodology proposed by Andersen and Petersen (1993).

Given the size of input-output matrices and the number of banks analyzed one can question the convergence of the results. The convergence issue can be solved by aggregating the input/output matrix. The fewer variables you have and the more observation, the more you tend to obtain convergent estimators of efficiency.

Daraio (2007) proposes a method to aggregate the variables, by representing an n-dimensional data cloud through a single vector that will hold maximum of information. She suggests normalizing data by dividing to dispersion or medium, as this has no effect on the obtained efficiency scores.

Also, Daraio points out that the weights forming the vector that retains the maximum of the total entropy are actually eigenvectors corresponding to the largest eigenvalues corresponding to $X^TX$ matrix (for input) and $Y^TY$ (for outputs).

If in principal component analysis eigenvalue report $\lambda_i/(\lambda_1+...+\lambda_n)$ represents the amount of information retained in the first eigenvalue from the total information, for aggregation, the same report is actually the inertia factor.

The following table (Table 1) shows the inertia and the cumulative inertia for input and output aggregation.

| Input Inertia | Cumulated Inertia | Output Inertia | Cumulated Inertia |
|---------------|-------------------|----------------|-------------------|
| 0.3500099     | 0.214397          | 0.3219827      | 0.237655          |
| 0.7932023     | 0.700269          | 0.0956264      | 0.304321          |
| 0.4893204     | 1                 | 0.9419037      | 1                 |

Using inertia input / output vector of weights I aggregated all the inputs / outputs in a single vector of inputs / outputs; now we can easily represent data in a two-dimensional space. In the below table (Table 2), it can be observed the high level of correlation between aggregated inputs / outputs and initial input / output - aggregation was performed successfully, one factor summarizing much of the information contained in the initial vectors of input / output.

| Fees & Commission Expenses | Interest Paid | Provisions Expenses | Aggregated Input |
|----------------------------|--------------|---------------------|------------------|
| Fees & Commission Expenses | 1            | 0.929535            | 1                |
| Interest Paid              | 0.929535     | 1                   | 1                |
| Provisions Expenses        | 0.514135     | 0.618522            | 1                |
| Aggregated Input           | 0.941371     | 0.997407            | 0.657195         |

| Net Income | Commissioning Revenue | Interest Revenue | Aggregated Output |
|------------|-----------------------|------------------|-------------------|
| Net Income | 1                     | 0.970988         | 1                 |
| Commission Revenue | 0.98682       | 0.986684         | 1                 |
| Interest Revenue | 0.989515       | 0.987249         | 0.9982            |
| Aggregated Output | 0.989515       | 0.987249         | 0.9982            |

Summarizing the initial vectors of output is even more successful than the aggregation of inputs, the aggregate output retaining much of the information contained in the original vectors.

Due to small number of observations (99 banks) and in order to ensure the relative convergence of the estimators obtained following the application of parametric methodology, I decided to work strictly with the aggregated vectors.
of inputs and outputs. I also used the logarithm function on this aggregated vectors and kept in the analysis all banks, including shikin.

The figure below (Fig 1) shows the frontier obtained through methodologies prior to SFA, namely regression (dashed line) and COLS (dotted line) as frontiers that cover the data, and also the SFA frontier (continuous line).

![Frontier representation](image1.png)

**Fig. 1. Frontier representation (from up-down) – COLS, SFA, regression.**

Next figure (Fig 2), shows the histogram associated to the efficiency scores obtained using COLS methodology. It can be observed a low percentage of perfect efficient banks, due to the effect of intercept alteration in order to ensure data coverage.

![Probability distribution of COLS estimators](image2.png)

**Fig. 2. Probability distribution of COLS estimators.**

The histogram displays the efficiency measures calculated according to Shepard distances. We can observe that most of the banks are found in the efficiency interval of 0.2-0.5. In fact, being given the input vector, the banks produce 20% to 50% of the output that can be produced.

Applying the methodology of likelihood function maximization, we obtain the following results (Fig 3):
From the data above, we can observe the significance of obtained statistical estimators, with an associated p-value t-test equal to 0. Also, given the term $\lambda = 2.3958$, we obtain $\frac{\lambda^2}{\lambda^2 + 1} = 0.8516$, which means that from the total variance, 85.16% is due to inefficiency, the remainder being due to noise term.

R program also makes it possible to calculate variances directly related to inefficiency term and error term. Thus, $\sigma^2 = 0.0647$ and $\sigma^2 = 0.3717$.

Also in Figure 1, the continuous line represents the SFA frontier, according with parametric methodology. It can be observed that some observations are above SFA frontier, due to the existence of the error term.

Using Equation 4 and R software I present further a histogram of SFA estimators (Fig 4).

The average of estimated efficiency measures through SFA is 0.657, also quite close to the average measures obtained using nonparametric DEA technique (a previous research led to an average level of 0.75) (Coelli, 2003; Farrell, 1957, Charnes, Cooper 1987).

In Table 3, I propose a comparison of estimators obtained using DEA, COLS and SFA.

| Parameters | Std.Err | t-value | Pr(>|t|) |
|------------|---------|---------|---------|
| Intercept  | 2.9248  | 0.0856  | 34.168  | 0       |
| x          | 0.8042  | 0.02944 | 27.317  | 0       |
| lambda     | 2.3958  | 0.68544 | 3.495   | 0       |
| sigma2     | 0.43654 |         |         |         |
| sigma2v    | 0.064785|         |         |         |
| sigma2u    | 0.3717695|        |         |         |
| log likelhood | -58.03895 |     |         |         |
| Convergence| 1       |         |         |         |

Fig. 3. Summary of SFA results.

Fig. 4. Distribution of SFA estimators.

Table 3. Correlation matrix of average estimates DEA, SFA and COLS

|        | DEA     | SFA    | COLS   |
|--------|---------|--------|--------|
| DEA    | 1       | 0.770639| 0.643006|
| SFA    | 0.770639| 1      | 0.915907|

In Table 3, I propose a comparison of estimators obtained using DEA, COLS and SFA.
It can be observed that the results obtained using DEA and SFA are relatively highly correlated. If during the DEA research I found approximately 25% from total banks, lying on the DEA frontier, for SFA, no firm is perfect efficient.

3. Conclusion

The research succeeded in finding the average score of efficiency associated to Japanese Banking System, using SFA as parametric technique used for estimating the efficiency of a decision making unit. Also, it was shown the correlation between SFA and other similar methods (COLS) and even a non-parametric technique (DEA).

Following research should ensure the significance of obtained parameters and the convergence of the obtained estimators.

References

Andersen P., Petersen N., 1993. A Procedure for Ranking Efficient Units in Data Envelopment Analysis, Management Science, 39:1261-1264.
Banker R.D., Charnes A., Cooper W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis Management Science 30, 1078–1092
Bogetoft P., Otto L., 2010. Benchmarking with DEA, SFA and R, Springer, New York
Charnes, A., Cooper, W.W., 1990. Polyhedral cone—ratio DEA models with an illustrative application to large commercial banks, Journal of Econometrics 46, 73–91
Coelli T., Prasada D.S., 2005. An introduction to efficiency and productivity analysis, Springer
Cooper W. W., Charnes A., Rhodes E., 1987. Measuring the Efficiency of decision making units
Drake L., Hall M., 2003. Efficiency in Japanese banking: An empirical analysis, Journal of Banking Finance, 891-917
Fare R., Grosskopf S., Lovell K.C.A., 1985. The Measurement of the Efficiency of Production, Kluwer Nijhoft, Boston.
Farell M.J. 1957. The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, 253-290
Fethi M., Pasiouras F., 2009. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey, European Journal of Operational Research, 189-198
Fukuyama, H., 1993. Technical and scale efficiency of Japanese commercial banks: A non-parametric approach, Applied Economics 25, 1101–1112.
Simar L., Cinzia D., 2007. Advanced robust and nonparametric methods in efficiency analysis. Methodology and Applications, Springer
Thanassoulis E., Kortelainen M, Allen R., 2012. Improving envelopment in Data Envelopment Analysis under variable returns to scale, European Journal of Operational Research, 175-185