2nd Assignment in Special Topics in Buisiness economics. USA Banks: Efficiency-productivity levels, inputs-outputs and ineficiency estimations.

January 25, 2023

Moustakas Andreas (1067656)
Department of Economics, Master of Science:
"Applied Economics and Data Analysis"

1 Abstract.

Within this report the efficiencies and productivity of US banks will be computed and displayed. Although the high levels of efficiency US banks have, we observe that there are some alternations between banks with different region. In addition, a Cobb-Douglas and a Translog production function have

been formed in order to estimate bank's production (output) given its liabilities, equity, deposits and salary (inputs). A statistically significant model confirmed that bank's output variance can be interpreted by input's variance. Lastl objective of this study was to examine the relationship between bank's inefficiency and capital adequancy, asset quality, return on assets and liquidity. With the ordinary least squares approach, the regression model could not estimate statistically significant coefficients, thus, the inefficiency changes can't be interpreted by bank's characteristic's changes.

2 Introduction.

There is a wide range of research and analysis on the efficiency of banks, as it is a vitally important branch for the economy of a country (Robert Lensink, Aljar Meesters, Ilko Naaborg, 2007). The sample used in this study is composed up of information from US banks. Information such as the bank's region, labour used, assets and more (detailed in the data section). This article's objective is to compute the efficiency levels of US banks, their Malmquist productivity index (MPI) and its components (scale, technical and efficiency change). Moreover, a Cobbb-Douglas and a Translog production function will be constructed. Also we will estimate the effect on bank's inefficiency of some key factors such as bank's capital adequacy, banks asset quality , bank's liquidity and more, the related section of the analysis identifies the relationship to more details. And the last objective of this analysis is to test the effect of Herfindahl index variable and the dummy variable which takes the value of 1 for commercial banks and 0 for savings banks. The financial system in the United States is stable, as evidenced by the fact that its bonds are a secure investment and inspire trust (Levy H, Lerman Z, 1988). So, it would be reasonable to assume that the US's banking system will have high efficiency and productivity levels. However, these levels may differ from a region to another. For example, in a wealthier geographical region the banks may have higher levels of efficiency and productivity in comparison with others. Furtermore two production functions (Cobb-Douglas and Translog)) will be constructed with the sfa method in order to compute the bank's output given its input. For these estimations the logarithms of inputs/outputs will be used, And last but not least, the amount of influence on inefficiency by its characteristics for each bank wil be estimated.

3 Data.

As far as the data are concerned, the data set comes from the Federal Deposit Insurance Corporation (FDIC) and refers to banks of United States. Having the data pretreatment done we have data for 2,199 US banks in hand, the data have also been balanced, thus a bank has its data for a period of 6 years (2011-2016), also shown in Table 1. These kind of panel data require lots of attention in order to get them in the appropriate condition. In this report, the banks have been clustered by their geographical district, Western (3,438 obs.), Southern (3,756 obs.), Northeastern (2,298 obs.), and Central (3,702 obs.) also shown in Table 2. The clustering of the banks was based on (Paul Krugman, 1998) since the geographic area constitutes a crucial characteristic for the growth of many markets as well as the development of the banks. Furthermore the emprical models (1),(2) data are depicted on Table 3, where summary statistics of the inputs/output have been presented. The models (1),(2) independent variables are the inputs of the banks, liabilities, equity, deposits and salary. These variables form the basic and significant factors for us to estimate the dependent variable. Which is the banks output, total production. As for the ordinary least squares method's equations (3) and(4), their independent variables are shown in Table 4, which are the capital adequacy, asset qualiy (30d and 90d), returns on assets, bank's liquidity, their Herfindahl index and wether their a commercial bank or a saving bank. The aforementioned characteristics composes the required data in order to estimate their inefficiency levels.

Table 1: Descriptive Statistics of Year.

Year	Frequency	Percent
2011	2,199	16.67
2012	2,199	16.67
2013	2,199	16.67
2014	2,199	16.67
2015	2,199	16.67
2016	2,199	16.67
Total	13,194	100

Source:https://www.fdic.gov

Table 2: Descriptive Statistics of District.

District	Frequency	Percent
Western	3,438	26.06
Southern	3,756	28.47
Northeastern	2,298	17.42
Central	3,702	28.06
Total	13,194	100

Source:https://www.fdic.gov

Table 3: Summary Statistics of Inputs/Output.

Variable	Obs	Mean	Std. Dev.	Min	Max
Total Production	13,194	859185.1	1331882	27356	1.66e+07
Liabilities	13,194	479179.5	730226.8	18569	8495643
Equity	13,194	57748.7	92439.8	2151	1404567
Deposits	13,194	125071.6	169097.6	1841	2727289
Salary	13,194	8205.4	12026.2	245	150645

Source:https://www.fdic.gov

Table 4: Summary Statistics of Independent variables (ols).

Variable	Obs	Mean	Std. Dev.	Min	Max
Capital adequacy	13,194	.3428	.2721	-1.0743	3.4138
Asset quality (30d)	13,194	.0050	.0060	0	.0800
Asset quality (90d)	13,194	.0007	.0038	0	.1655
Return on assets	13,194	.9153	.5951	-3.2963	3.3901
Liquidity	13,194	.3021	.1430	.0154	.9353
Herfindahl index	13,194	.0338	.0498	.0058	1
Commercial	13,194	.8863	.1136	0	1
Saving	13,194	.1136	.3174	0	1

Source:https://www.fdic.gov

Vrs, Crs and Scale efficencies.

In this section the vrs, crs and scale efficiencies are depicted on Table 1-6. Specifically with tha data envelopment analysis (dea) method, the efficients of every bank was computed. Dea is a mathematical programming method for evaluating the relative efficiency of decision making unit (dmu) with multiple inputs and outputs. With this method the effieciencies of every bank for every year were calculated under crs and vrs assumptions. The following tables presents summary statistics of the bank's efficiencies for through time by district.

Western:

Over the years (2012-2016), consistent degree of high levels of efficiency is seen with minimal variations in the western region.

Table 5: Efficiencies of Western US Banks.

Western

Eff.	Mean	Median	Max	Min	
Vrs	0.87	0.87	1	0.69	
Crs	0.86	0.85	1	0.57	2012
Scale	0.97	0.99	1	0.65	
Vrs	0.87	0.87	1	0.57	
Crs	0.86	0.86	1	0.61	2013
Scale	0.98	0.99	1	0.61	
Vrs	0.88	0.88	1	0.55	
Crs	0.86	0.87	1	0.50	2014
Scale	0.98	0.99	1	0.66	
Vrs	0.89	0.90	1	0.57	
Crs	0.88	0.87	1	0.53	2015
Scale	0.97	0.99	1	0.70	
Vrs	0.89	0.90	1	0.58	
Crs	0.88	0.88	1	0.53	2016
Scale	0.97	0.98	1	0.69	

Source:https://www.fdic.gov

Southern:

In southern region the efficiences are relatively stable, the minimum efficiencies are quite lower than western's.

Table 6: Efficiencies of Southern US Banks.

Southern

Eff.	Mean	Median	Max	Min	
Vrs	0.86	0.86	1	0.65	
Crs	0.84	0.84	1	0.58	2012
Scale	0.97	0.98	1	0.58	
Vrs	0.87	0.88	1	0.63	
Crs	0.85	0.85	1	0.57	2013
Scale	0.97	0.99	1	0.57	
Vrs	0.86	0.87	1	0.63	
Crs	0.84	0.85	1	0.58	2014
Scale	0.97	0.99	1	0.58	
Vrs	0.87	0.87	1	0.63	
Crs	0.85	0.85	1	0.57	2015
Scale	0.97	0.98	1	0.57	
Vrs	0.87	0.88	1	0.64	
Crs	0.85	0.86	1	0.57	2016
Scale	0.97	0.99	1	0.57	

Source:https://www.fdic.gov

Northeastern:

Northeastern district also has some little variations through time, but the minimum efficiencies are the highest of the region's group. Seems to be in a better place, as far as efficiencies are concerned.

Table 7: Efficiencies of Northeastern US Banks.
Northeastern

Eff.	Mean	Median	Max	Min	
Vrs	0.89	0.89	1	0.68	
Crs	0.88	0.88	1	0.67	2012
Scale	0.98	0.99	1	0.76	
Vrs	0.90	0.90	1	0.66	
Crs	0.89	0.89	1	0.62	2013
Scale	0.98	0.99	1	0.78	
Vrs	0.91	0.91	1	0.68	
Crs	0.89	0.90	1	0.64	2014
Scale	0.98	0.99	1	0.72	
Vrs	0.91	0.91	1	0.71	
Crs	0.89	0.90	1	0.60	2015
Scale	0.98	0.99	1	0.70	
Vrs	0.92	0.92	1	0.72	
Crs	0.91	0.92	1	0.60	2016
Scale	0.98	0.99	1	0.71	

Source:https://www.fdic.gov

Central:

Lastly , central region makes no exception, as there are not much of a differences on the efficiency levels.

Table 8: Efficiencies of Central US Banks.
Central

Eff.	Mean	Median	Max	Min	
Vrs	0.87	0.87	1	0.64	
Crs	0.85	0.85	1	0.63	2012
Scale	0.97	0.98	1	0.66	
Vrs	0.87	0.87	1	0.60	
Crs	0.86	0.86	1	0.58	2013
Scale	0.98	0.99	1	0.63	
Vrs	0.88	0.88	1	0.64	
Crs	0.86	0.86	1	0.64	2014
Scale	0.97	0.98	1	0.64	
Vrs	0.88	0.88	1	0.61	
Crs	0.87	0.87	1	0.61	2015
Scale	0.98	0.99	1	0.68	
Vrs	0.89	0.89	1	0.58	
Crs	0.87	0.88	1	0.58	2016
Scale	0.98	0.99	1	0.66	

Source:https://www.fdic.gov

Having the summary statistics depicted we express that, with the passing of years, there have not been much of changes on the efficiency levels of US banks. As we first stated, US have a solid economical system, so there is no wonder why the efficiencies does not fluctuate. Howover, we observed that Northeastern region has the highest levels of efficiencies. It seems like banks efficiencies have a relationship with their region's wealth, thus Northeastern with the highest levels of revenues have the highest levels of bank's efficiencies.

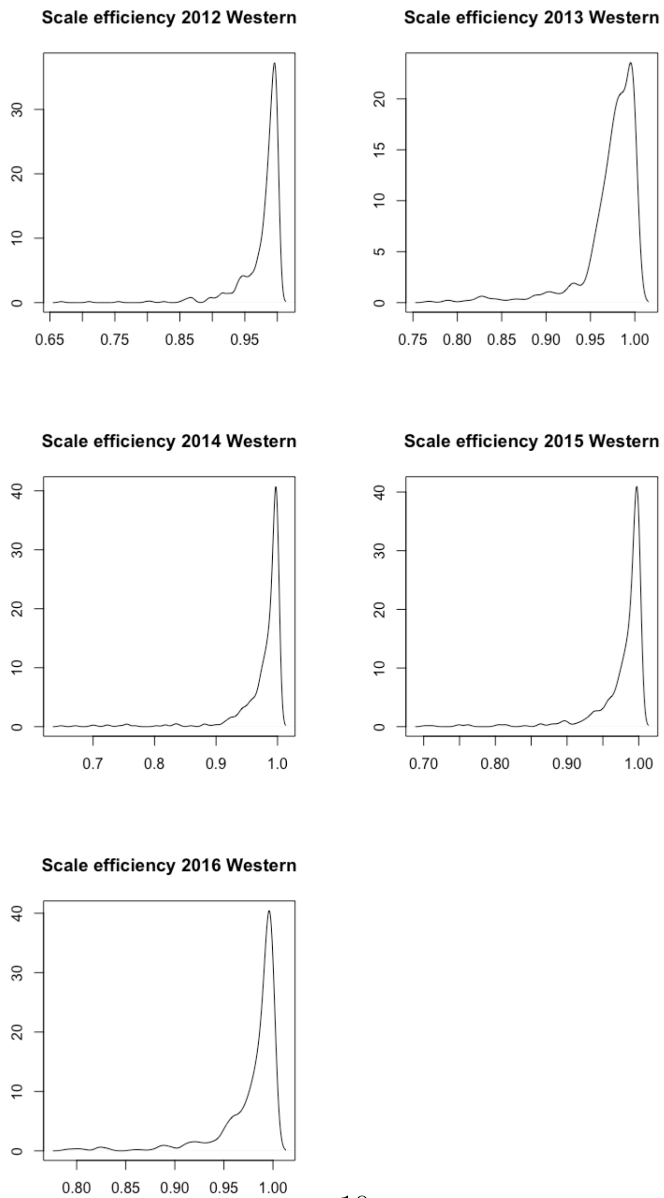
Distribution of Scale efficiency

Kernel density graphs will be presented on Figures 1-4. In these graphs we can observe the tendency of the scale efficiencies of every bank's accrued year 2012-2016. A density plot shows us how a numerical variable's distribution looks like. It displays the probability density function of the variable using a kernel density estimate. It is used in the same way as the histogram but has been smoothed out. In bibliography, kernel density graphs are frequently used to present distributions, thus this study is no exception. As expected, the levels of the scale efficiency of banks are significantly high (Figures 1-4), as they represent a strong banking system . There have been noted some fluctuations with the passing of time. In a broad sense, bigger revenue's regions have also a precedence on efficiency levels.

Western:

For the western banks we see that the density curves have negative skewness (left skewed distribution) ,consequently, the median exceeds the mean. In general terms, the distribution does not change significantly over time.

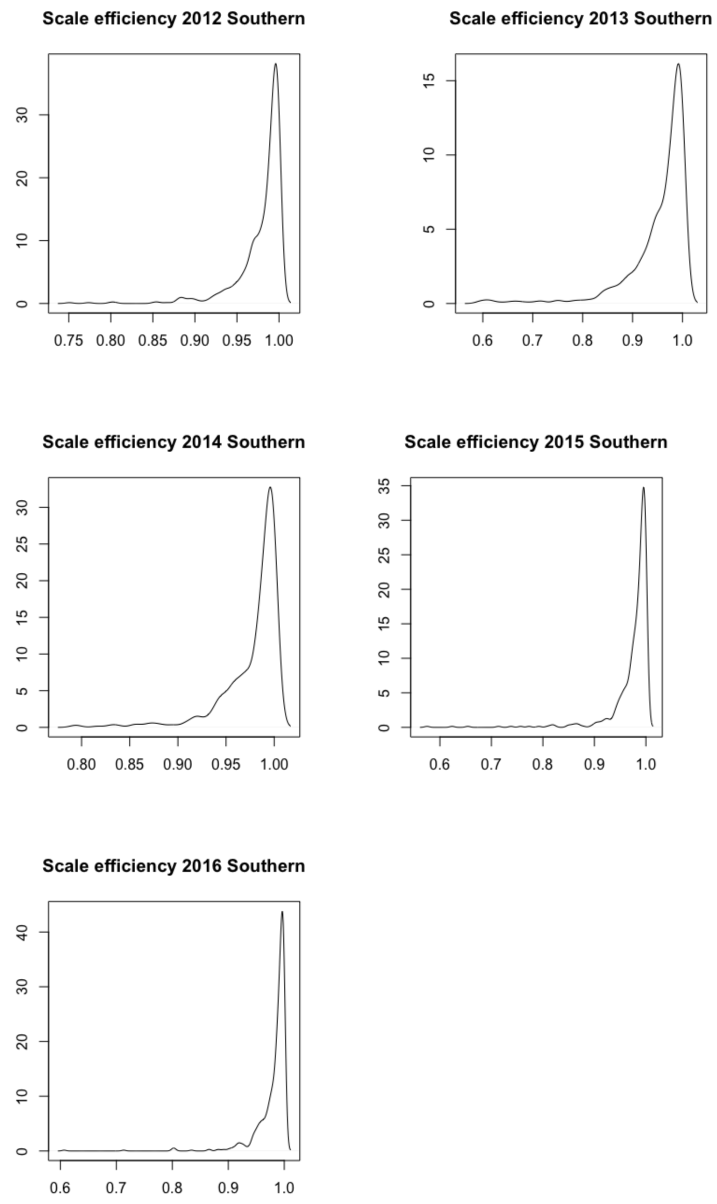
Figure 1: Kernel density graph of Scale efficients (Western 2012-2016)



Southern:

As far as the Southern banks are concerned, the density curves have also negative skewness, its mean is lesser than the median.

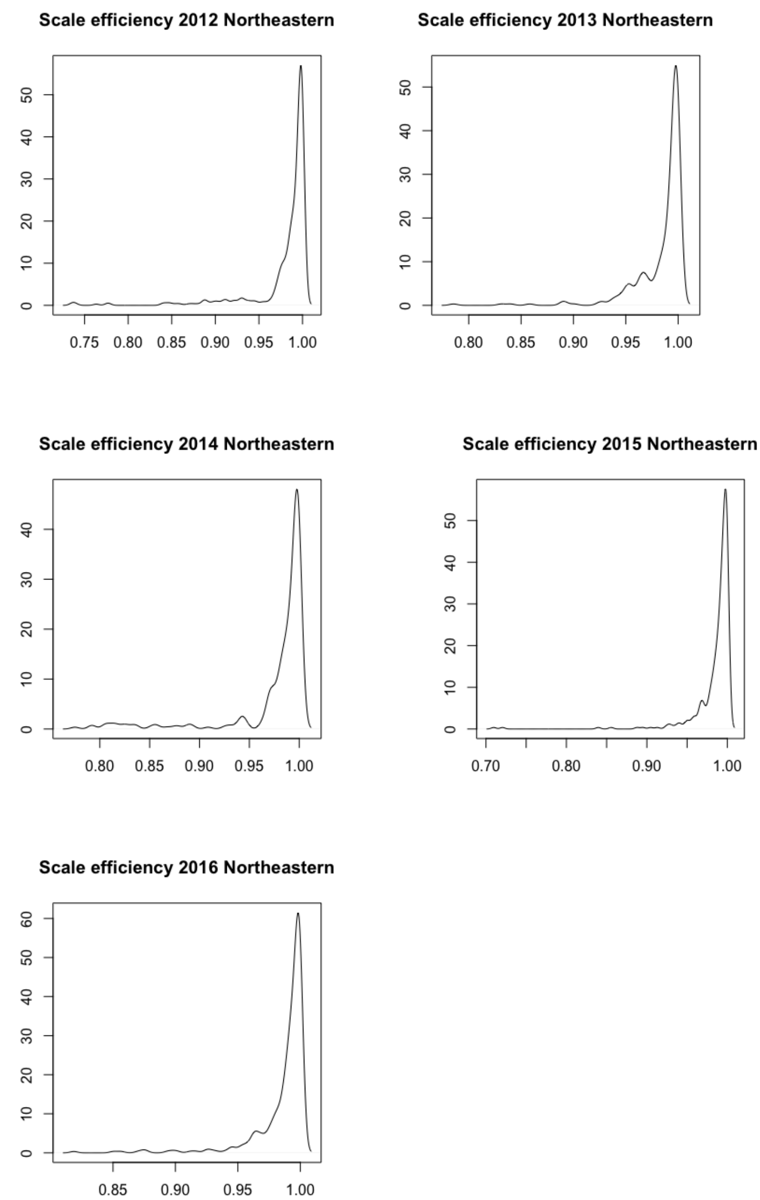
Figure 2: Kernel density graph of Scale efficients (Southern 2012-2016)



Northeastern:

In the Southern banks, the negative skewness of the density curves are observed, as well as that the efficiencies have a little alternations as years past. Also the median is higher than mean due to the left skewed distribution.

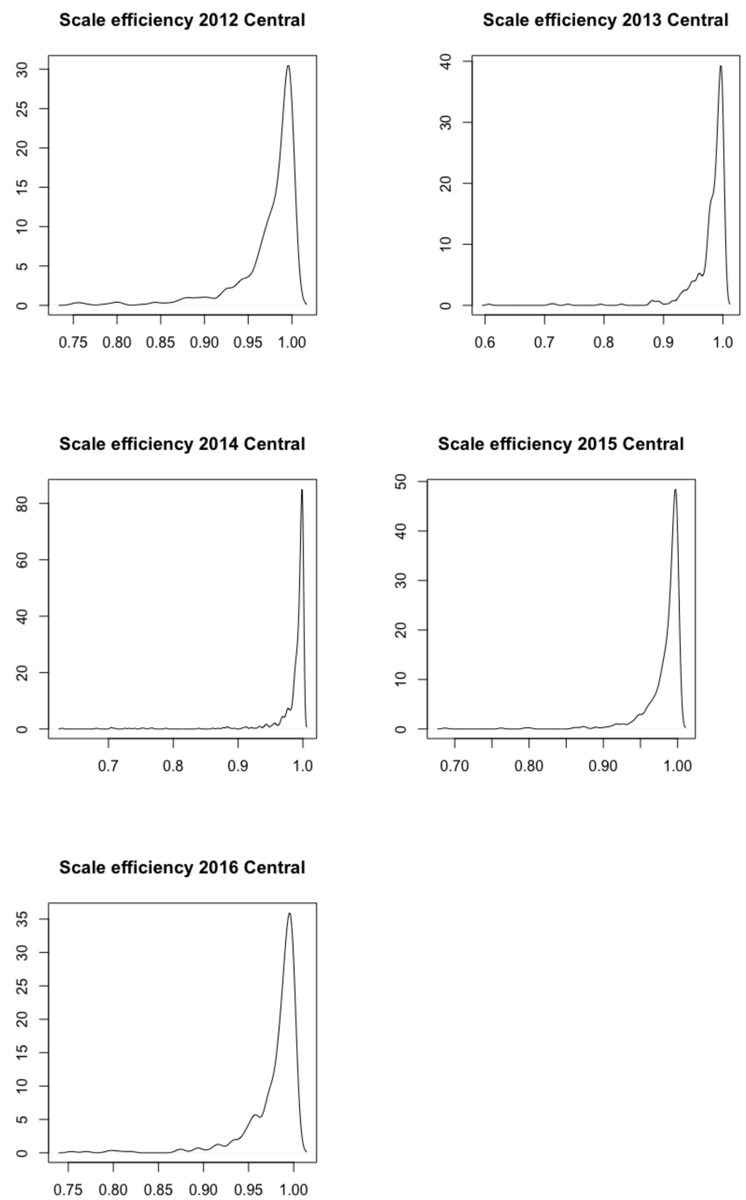
Figure 3: Kernel density graph of Scale efficients (Northeastern 2012-2016)



Central:

Finally, we see that the skewness of the density curves coreresponds to the rest of the districts. It has negative skewness thus the median is higher than mean.

Figure 4: Kernel density graph of Scale efficients (Central 2012-2016)



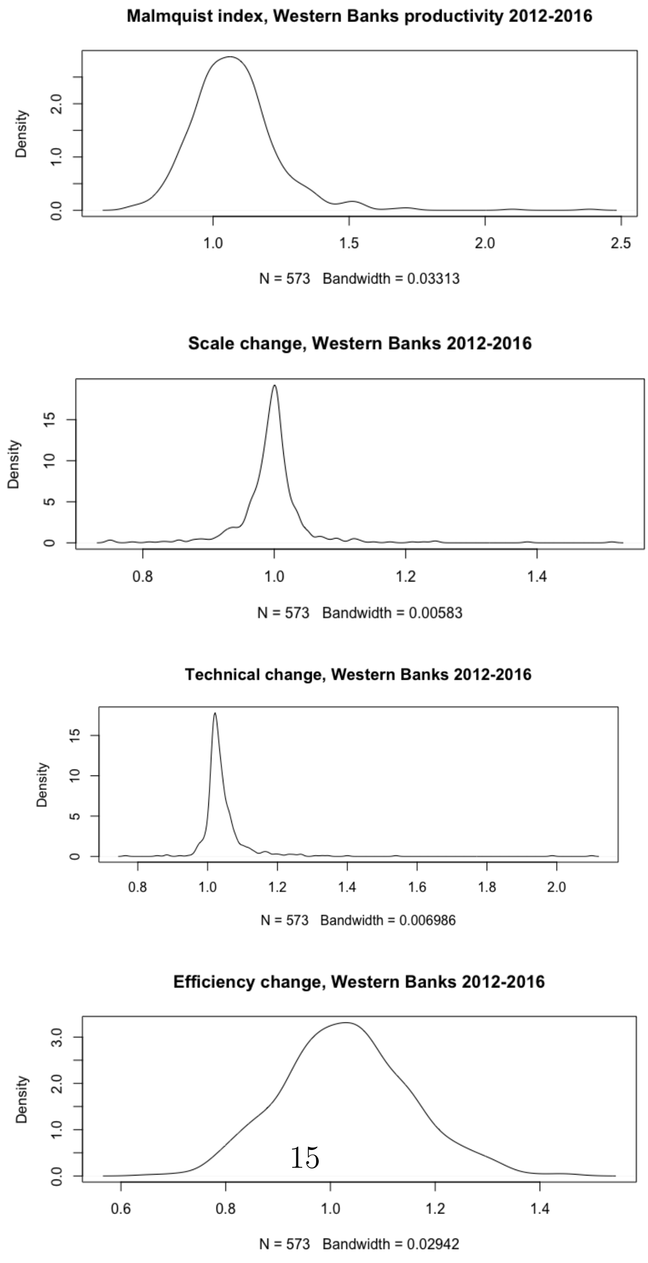
Mamquist index, Scale change, Technical change and Efficiency change.

The Malmquist productivity index (MPI) has been calculated in order to look into bank's productivity by district. Furthmore, it was significant to compute its productivity components, such as, scale change, technical change and effciency change so as to examine the differences in producticity. As shown on Figures 5-8, there are some alternations on MPI and its components by geographic location. In favor of the wealthier districts, productivity sligthly differs. Northeastern MPI and its components seems to have more banks on the highest standards of producticity. On the below Figures 5-8, these measures have been presented with density graphs. As stated above density graphs helps us to understand the distribution of the values and some of their characteristics.

Western:

In western banks, the malmquist productivity index has its density curves with positive skewness (right skewed), thus the mean is higher than the median. Also the scale and technical change has right skewd curves, whereas, the efficiency change seems to be a Gaussian distribution (distributed normaly) with its mean equal to median (mean=median=1).

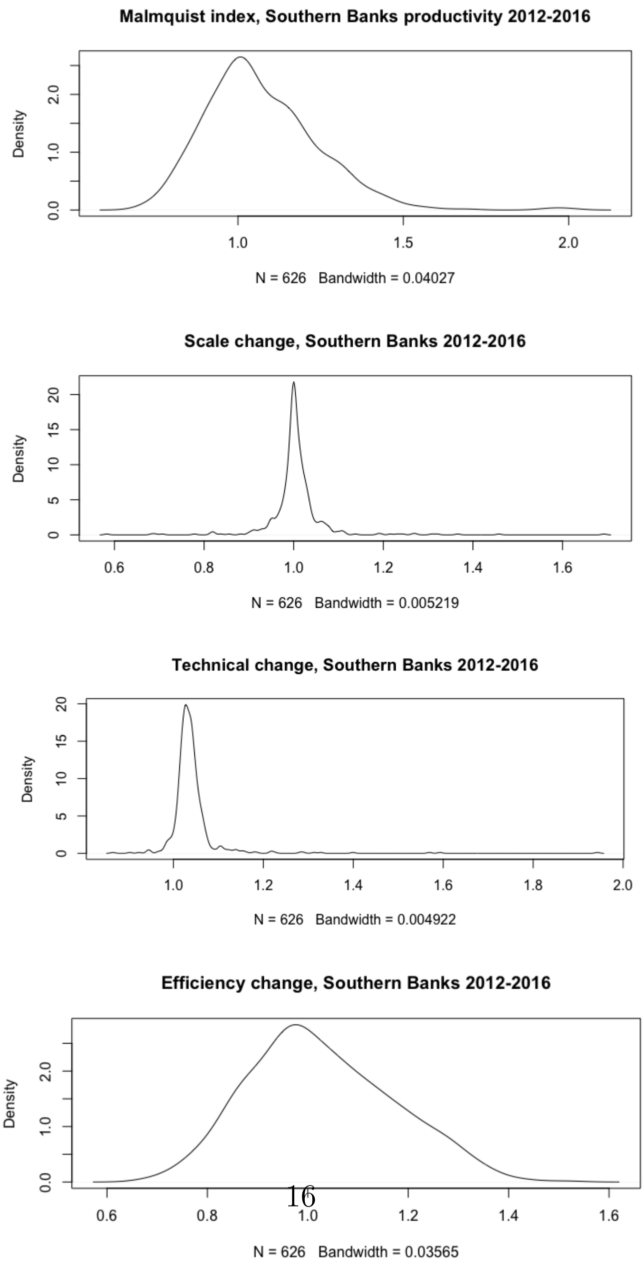
Figure 5: Kernel density graph of MPI - Components



Southern:

Southern banks are consisted of right skewed MPI , scale change and technical change, with their means to be higher than their medians. The efficiency change follows the normal distribution, with the lack of skeweness median is equal to mean.

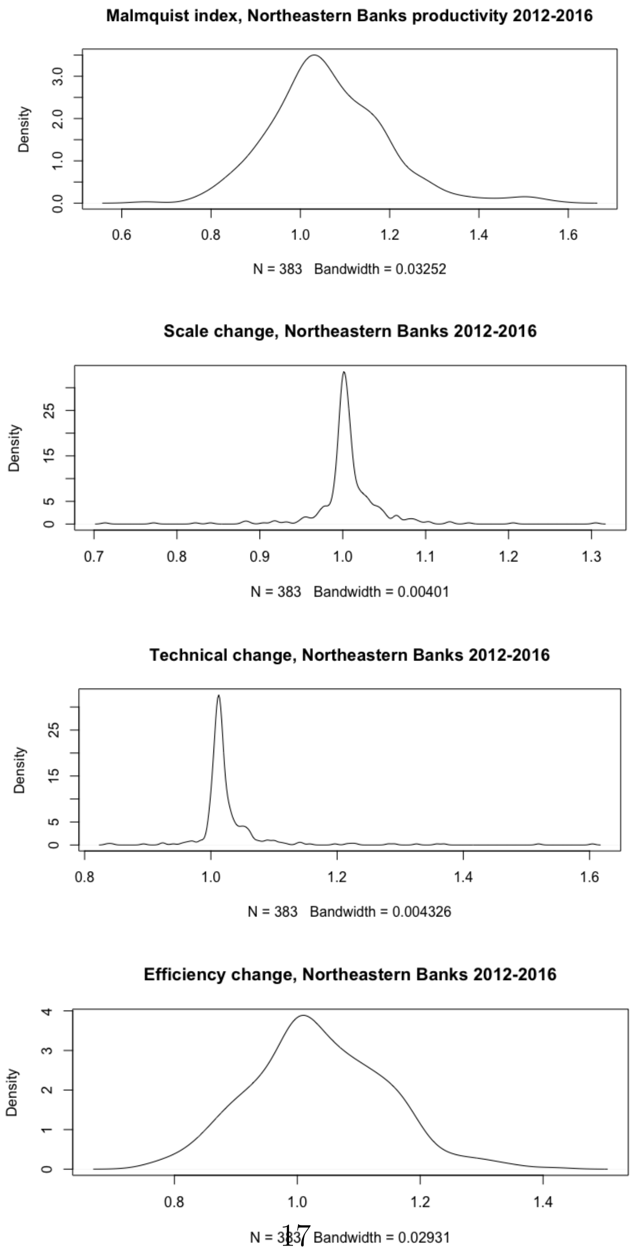
Figure 6: Kernel density graph of MPI - Components



Northeastern:

As for the northeastern banks MPI, scale change and efficiency change are composed by a Gaussian distribution. Their distributions are symmetric to 1. Technical change has a density curve with right skewness.

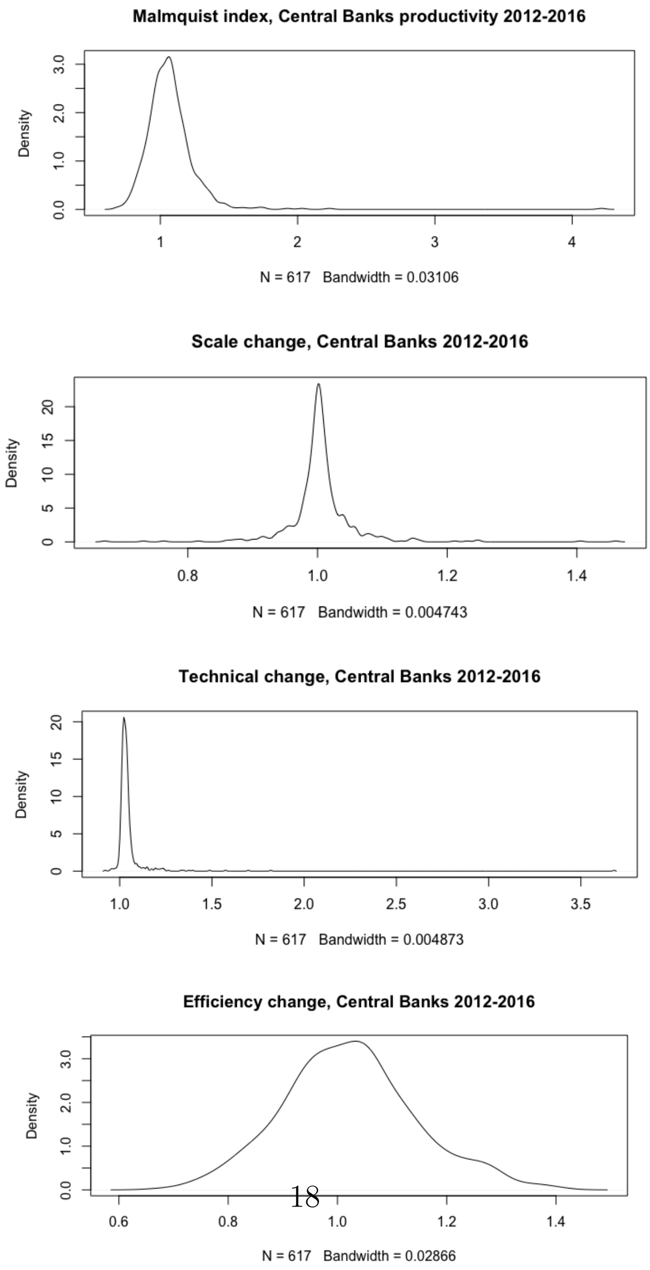
Figure 7: Kernel density graph of MPI - Components



Central:

In central banks malmquist productivity index follows a right skewed distribution, as well as the technical change (mean¿median). Both scale changes and efficiency changes follow the normal distribution which is symmetric to 1 (mean=median=1).

Figure 8: Kernel density graph of MPI - Components



4 Empirical model.

In order to estiamte banks total production, two models have been constructed. Firstly, a Cobb-Douglas production function is set (equation 1), given the inputs/output of each bank. With this function we are capable to determine input to output ratios for efficient production and to estimate technological change in production methods. As for the latter approach, a translog production function is structured (equation 2), this method is very commonly used, as it is a generalisation of the Cobb-Douglas function,.

Cobb-Douglas production function.

$$Y_i = L_i^a E_i^b D_i^c S_i^d (1)$$

Where:
Y: Total production.
L: Liabilities.
E: Equity.
D: Deposits.
S: Salary.
i: Bank (1 - 2,199).
a, b, c, d: Output elasticities.

Translog production function.

$$lnY_i = a_0 + \sum_{n=1}^{4} a_{i,n} lnX_{i,n} + 1/2 \sum_{n=1}^{4} \sum_{m=1}^{4} a_{i,n,m} lnX_{i,n} lnX_{i,m} + u_i (2)$$

Where:
Y: Total production.
X: Set of independent variables.
(Liabilities,Equity,Deposits,Salary).
n: Independent variable (1 - 4).
a: Inercept (n=0).
a: Coefficient (n= 1 - 4).
i: Bank (1 - 2,199).
u: Residuals.

Bank's inefficiency estimation. (OLS)

It is very vital to estimate the inefficiencies of banks in order to be able to improve their efficiencies. If a firm is able to locate its weaknesses, has better chances in improving them. These empirical models (equation 3, 4) have been estimated with ordinary least squares method, as it is common technique for estimating coefficients of linear regression equations which describe the relationship between independent variables and a dependent variable. In this case, equation 3 uses a set of independent variables (presented in Table 4) which are needed in order to interpret the inefficiency variation.

$$I_i = a_0 + a_1 C_i + a_2 N3_i + a_3 N9_i + a_4 R_i + a_5 L_i + u_i (3)$$

$$I_i = a_0 + a_1 C_i + a_2 N3_i + a_3 N9_i + a_4 R_i + a_5 L_i + a_5 H_i + a_6 Com_i + u_i (4)$$

Where:
I: Inefficiency.
C: Capital adequacy.
N3: Asset quality (30 through 89 days).
N9: Asset quality (90 or more days).
R: Return on assets.
L: Bank's liquidity.
H: Herfindahl index.
Com: Type of bank (Comercial, dummy).
u: Residuals.

5 Results.

In this section the results of the previous models and methods will be presented and analysed. The Cobb-Douglas and translog production function were capable to estimate the right models which are statistically significant. Input's level of influence on the outputs are displayed on Tables 9 and 10. As far as the estimation of bank's inefficiencies are concerned, the results were not ideal, wheras the coefficients were not statistically significant. Thus their relationship could not be interpreted

Cobb-Douglas production function (Results).

Having the results of the Cobb-Douglas production function, we can observe that every given input is statistical significantly on the strictest significant level. From a general perspective this means that the levels of bank's production (output) is strongly correlated with these inputs (Table 3, Table 9). Thus, bank's Liabilities, equity, deposits, and salary's variation can affect the production's variation.

Table 9: Cobb-Douglas production function.

	Estimate	Std. Error	z value	Pr(>|z|)	
Intercept	0.9601	0.0173	55.2267	<2e-16	***
log(L)	0.8201	0.0037	218.3291	<2e-16	***
log(E)	0.1349	0.0027	49.3835	<2e-16	***
log(D)	0.0103	0.0016	6.1852	6.2e-10	***
log(S)	0.0579	0.0025	22.8738	<2e-16	***
sigmaSq	0.0247	0.0006	40.2408	<2e-16	***
gamma	0.9590	0.0011	846.3120	<2e-16	***
time	0.0385	0.0011	32.8145	<2e-16	***

Signif. codes: '***' 1% , '**' 5% , '*' 10%

Translog production function (Results).

With the translog function, we also see that inputs are statistically significant (Table 10). Having them calculated, we are able to look into the level of influence of the output by the inputs. There is a substantial correlation between the production levels of banks and bank's liabilities, equity, deposits, and salary.

Table 10: Translog production function.

	Estimate	Std. Error	Z value	Pr(>|z|)	
Intercept	3.8226e-01	1.6041e-01	2.3830	0.0171737	*
log(L)	8.5698e-01	5.3561e-02	16.0001	<2.2e-16	***
log(E)	1.3104e-01	4.2989e-02	3.0482	0.0023024	**
log(D)	6.8624e-02	2.4379e-02	2.8149	0.0048792	**
log(S)	5.4368e-02	2.8330e-03	19.1910	<2.2e-16	***
I(0.5 * (S) * (S))	2.1991e-12	5.6135e-12	0.3918	0.6952417	
I(0.5 * log(L)^2)	-1.6015e-02	1.2244e-02	-1.3080	0.1908651	
I(0.5 * log(X)^2)	4.2044e-03	7.2146e-03	0.5828	0.5600479	
I(0.5 * log(D)^2	-1.1495e-03	2.3342e-03	-0.4924	0.6224064	
I(log(L) * log(E))	1.1076e-02	8.7271e-03	1.2691	0.2044029	
I(log(L) * log(D))	6.4798e-03	4.5468e-03	1.4251	0.1541216	
I(log(E) * log(D))	-1.4643e-02	4.0399e-03	-3.6246	0.0002894	***
I((S) * log(D))	-2.4985e-07	1.5239e-07	-1.6396	0.1010902	
I((S) * log(E))	-6.1026e-07	3.1835e-07	-1.9169	0.0552469	.
I((S) * log(L))	7.3522e-07	2.8188e-07	2.6083	0.0091001	**
sigmaSq	3.2676e-02	1.0298e-03	31.7317	<2.2e-16	***
gamma	9.6690e-01	1.2190e-03	793.1954	<2.2e-16	***
sigmaSqU	3.1594e-02	1.0328e-03	30.5920	<2.2e-16	***
sigmaSqV	1.0816e-03	1.5067e-05	71.7899	<2.2e-16	***
sigma	1.8077e-01	2.8483e-03	63.4634	<2.2e-16	***
sigmaU	1.7775e-01	2.9051e-03	61.1841	<2.2e-16	***
sigmaV	3.2888e-02	2.2906e-04	143.5798	<2.2e-16	***
lambdaSq	2.9210e+01	1.1125e+00	26.2563	<2.2e-16	***
lambda	5.4046e+00	1.0292e-01	52.5126	<2.2e-16	***
varU	1.1481e-02	NA	NA	NA	
sdU	1.0715e-01	NA	NA	NA	
gammaVar	9.1390e-01	NA	NA	NA	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Bank's inefficiency estimation. (OLS Results)

For the number 3 model, the independent variables affect the dependent by the measures shown in Table 11. However this relationship does not seems to be statistically significant on any level. Thus, these characteristics (Table 4) cannot interpret the bank's inefficiency. The constant term,is indeed statistically significant and bank's inefficiency is formed by its measure.

Table 11: Bank's inefficiency estimation. (Model 3)

	Estimate	Std. Error	T value	Pr(>|t|)	
Intercept	0.111518	0.004638	24.047	<2e-16	***
Capital adequacy	-0.001696	0.009713	-175	861	
Asset quality (30d)	0.465449	0.296753	1.568	117	
Asset quality (90d)	-0.415452	0.340946	-1.219	223	
Returns on assets	0.002814	0.004833	582	560	
Bank's liquidity	-0.002181	0.010660	-205	838	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The results of model 4 arise similar to the model's 3, the relationship is not a statistically significant one, thus, we draw the same conclusion. Having the Herfindahl index and the dummies about the bank's type (Commercial-Saving) does not help the model to estimate the right coefficients.

Table 12: Bank's inefficiency estimation. (Model 4)

	Estimate	Std. Error	T value	Pr(>|t|)	
Intercept	0.111345	0.005870	18.969	<2e-16	***
Capital adequacy	-0.001400	0.009724	-144	886	
Asset quality (30d)	0.468552	0.296942	1.578	115	
Asset quality (90d)	-0.411287	0.341161	-1.206	228	
Returns on assets	0.002862	0.004861	589	556	
Bank's liquidity	-0.001758	0.010807	-163	871	
Herfindahl index	0.019795	0.027446	721	471	
Commercial dummy	-0.000950	0.004890	-194	846	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6 Conclusion.

The data set's clustering was crucial to this report, as the main objective was to examine the bank's efficiency levels by their geographical region. Had our sample prepared, we were able to calculate the efficiencies and productivity of each bank on given year. The results confirmed the initial hypothesis that bank's district composes a key charasteristic of its performane. Moreover, with the sfa method, two production functions were constructed with the given variables. Having Cobb-Douglas (equation 1) and a translog (equation 2) production function, the production levels (outputs) were interpreted by the bank's inputs. Finally, for the inefficiencies estimation, the ordinary least squares method (equation 3, 4) could not provide us with statistically significant coefficients. For that reason, the Herfindahl index was added to the model as well as the type of bank, but the regression model was still unable to estimate the relationship of bank's inefficiency with its characteristics.

7 References.

Lensink, R., Meesters, A., Naaborg, I. (2008). Bank efficiency and foreign ownership: Do good institutions matter?. Journal of Banking and Finance, 32(5), 834-844.

Levy, H., Lerman, Z. (1988). The benefits of international diversification in bonds. Financial Analysts Journal, 44(5), 56-64.

Krugman, P. (1998). What's new about the new economic geography?. Oxford review of economic policy, 14(2), 7-17.

Andriakopoulos, K., Kounetas, K. (2022). Large lending and banks performance. Is there any relationship? Empirical evidence from US banks. Bulletin of Economic Research.

Berger, A. N., Humphrey, D. B. (1991). The dominance of inefficiencies over scale and product mix economies in banking. Journal of Monetary Economics, 28(1), 117-148.

Chortareas, G. E., Girardone, C., Ventouri, A. (2012). Bank supervision, regulation, and efficiency: Evidence from the European Union. Journal of financial stability, 8(4), 292-302.