A geographically weighted approach to measuring efficiency in panel data: The case of US saving banks

This version: April 28, 2013

Abstract

This paper discusses a new approach to controlling for the environment when estimating efficiency. In response to the literature on the international comparison of bank efficiency, we draw the attention to a local dimension of comparison. By introducing geographical weights and estimating local frontiers for each US savings bank for the 2001-09 period, we find that the bank technical performance is higher for most banks in comparison to a fixed-effects approach. This result highlights the importance of taking into account the local environment and constraints while analyzing banks' performance, so as not to consider the factors that are exogenous to these institutions as inefficiencies. Further analysis could improve the weighs calculation by employing other measures of interconnectedness besides geographical distance.

Keywords: Stochastic Frontier, Environmental Variables, Bank Performance, Geographical Weights

JEL classification: G21, G28

---

1. Introduction

The estimation of bank efficiency is a recurrent subject of analysis in the literature (Lensink et al., 2008; Berger et al., 2009; Lozano-Vivas and Pasiouras, 2010). Bank efficiency reflects the efficiency of financial intermediation and, thus, has direct implications on social welfare. The literature has been developing several methods for the estimation of bank efficiency in a particular banking industry[1]. The availability of sound and reliable indicators plays an essential role in financial regulation. This paper proposes a new method to estimate the technical efficiency and we apply this method for US Saving Banks over the 2001-2009 period. We basically employ geographical weights in the stochastic frontier estimation so as to give more importance to neighboring banks in the calculation of bank efficiency.

The stochastic frontier analysis (SFA), proposed and developed by Aigner et al. (1977) and Meeusen and van den Broeck (1977), is a parametric approach that estimates a frontier for a set of banking systems and compares each bank in the sample to the frontier. Inefficiency is how distance this bank is from the frontier. Further improvements to the model were made in Battese and Coelli (1992) that specified time varying inefficiency scores and in Battese and Coelli (1995) that permitted the model to account for factors that influence both technology and inefficiency.

One interesting conclusion of the bank efficiency literature is that environmental conditions play a significant role in determining bank performance. According to both Lozano-Vivas et al. (2002) and Hasan et al. (2009), even though the European financial systems have become more integrated with the establishment of the European Monetary Union (EMU), there are still many differences in the regulatory and economic conditions among them. In other words, a comparison of the banks operating in different countries against a single reference could consider specific characteristics that a particular banking system is subject to as inefficiencies rather than a reflection of whether its management of resources is effec-

[1] See Berger and Humphrey (1997) for a survey on bank efficiency estimation methods. The paper documents that the different estimation methods yield inconsistent bank-specific efficiency scores, even though the average efficiency for an industry remains similar.

tive. Dietsch and Lozano-Vivas (2000) indicate how different results can be if one controls for the environmental conditions in the bank home country in relation to an uncontrolled specification. Therefore, they state that the estimation of a single frontier for heterogeneous banking markets without controlling for environmental variables may result in biased efficiency scores.[2]

Notwithstanding these methods, we propose a new model that is suitable to explicitly modeling environmental factors in the estimation of technical efficiency. We assume that banks that are close geographically to one another are subject to similar constraints[3]. This approach is the geographically weighted stochastic frontier (GWSF), where we estimate the frontiers for each bank in the sample (local frontiers). In each estimation, we consider a different bank as the benchmark and a weight is given to the other banks depending on their distance to this reference. We therefore implicitly control for the geographical factors that influence the efficiency of banks that are close to one another. An additional advantage of the GWSF is that we are able to employ it even within a country, in this case, the US. Even though other papers apply this geographical method, such as Samaha and Kamakura (2008) for the real estate market, we are the first to employ it in panel data.

The influence of geographical factors on banks' performance is an increasingly recurrent theme in the recent literature. **For instance, Bos and Kool (2006) regress inefficiency scores on a set of variables that reflect, among other things, the local economic environment of the bank and find that these factors do explain part of the inefficiency scores, even though to a limited extent. Also, Pasiouras et al. (2011) concludes that factors that are external to managerial control influence**

[2]Bos and Schmiedel (2007) propose another modification to the SFA in order to make the efficiency scores comparable among countries. They use meta-frontiers, which allows for heterogeneous technology among banks from different countries. The authors affirm that part of the inefficiency of a single frontier estimation might be due to the technology gap among countries.

[3]The literature has interest in discussing whether recent technological developments, such as the more frequent use of internet and mobile banking, have reduced the importance of the physical location where a bank operates. Even though the internet plays an increasingly important role in reducing the cost of distance (Berger, 2003), Degryse and Ongena (2004) reaffirm the importance of the geographical distance in lending relationships.

Greek cooperative banks' efficiency measured by a Data-Envelopment Approach (DEA). In this fashion, we will show that there could be a significant bias in the efficiency scores if one does not take into account the geographical characteristics where each bank (or branch) operates. Therefore, we contribute to the issue of comparing the performance of the banks that are subject to different environmental constraints by stating that they can differ significantly even within the same banking system. Some of these factors are observable, such as the size of the market, the different laws and regulations, and the accessibility of banking services to the population; other factors are unobservable. This method takes into account both types because it estimates the efficiency of a bank in comparison to its neighbors.

As a proof of the statement above, there is extensive evidence that US banks' performance is geographically dependent. Akhigbe and McNulty (2003) find that US commercial banks operating in metropolitan areas (MSA) have different efficiency levels than those in non-metropolitan areas for the years 1990-96. In fact, banks in an MSA are less profit efficient than those in the non-MSA. In addition, according to Tirtiroglu et al. (2011), bank productivity in the US appears to be geographically dependent among states, where the performance in one state is positively correlated with the performance of its neighbors. Finally, Berger and DeYoung (2001) note that the return on assets varies considerably with the region. These facts are a clear motivation for our exercise, where we apply this new method to US saving banks.

In addition, the US banking system presents other interesting features regarding the geographical field. First, not only are these banks subject to federal regulation, but they must also respond to state laws, which can exert different influences on the banking operation. Second, as DeYoung et al. (2004) state[4], the removal of the geographical restrictions that

[4]DeYoung et al. (2004) considers these changes in economic conditions and explores whether the effects in geography changed with banks' headquarters locations, the bank branch office locations and the bank depositor locations. They found that (1) mergers and acquisitions have allowed banks to move their headquarters from smaller to larger cities, (2) bank branches have moved farther away from headquarters and (3) the spatial density of deposits in the 50 largest metropolitan areas has remained remarkably stable because the commercial banking industry became more spatially concentrated during the 1990s, which is

were put in place with the McFadden Act of 1924 have allowed banks to operate across state lines and to acquire banks anywhere in the country, converting some subsidiaries and removing branching restrictions. The Riegle-Neal Act of 1994 led to a geographical expansion into new markets, where merger activities became more accepted by the banking industry. This merger process has increased and has improved the bank's ability to lend and monitor these loans from offices far away from headquarters. In fact, between 1980 and 1990, a period of consolidation and restructuring, many banks were taken over by other depository institutions to raise efficiency.

There is no denying that the study of US saving banks' efficiency fits our model because these banks have a stronger regional focus of operation than regular commercial banks. In other words, they tend to lend more to the institutions and enterprises that are close to where they are located. US saving banks tend to compete with others that operate in the same geographical location, as well. It is less likely that the more distant banks can affect how a specific saving bank performs. Also, these banks are of the utmost importance because they supply a large fraction of small and medium firms' demand for loans. It is clearly in the interest of bank's regulators to know exactly how these banks perform, thereby choosing the proper set of regulations for them.

The case of commercial banks is more complex. Since these banks tend to operate on a national-wide basis, local factors might not influence them over and above the national-wide factors. Having access to branch-level data, one could employ our method to estimate branch-specific efficiencies that controls for the region they are located. This way, it would be possible to compare the results from previous papers on branch efficiency (Berger et al., 1997; Paradi and Schaffnit, 2004; Portela and Thanassoulis, 2007; Paradi et al., 2011) On the other hand, the weights calculation might be generalized to different measures other than geographical distance. One suggestion may be the difference in the loan

evidence of gradual urbanization. The results suggest that the spatial distribution of deposits remained similar across time. The results also suggest that new technologies increase the ability of banks to manage credit relationships.

portfolio sectoral composition. Banks that lend to similar industries might be subject to common shocks originated from these sectors that are not necessarily related to managerial efficiency. We leave these questions for future research.

We structure the remainder of the paper as follows. Section 2 presents our methodology, where we define the GWSF model and all of the steps to estimate it. In Section 3, we present and summarize the data sources. In addition, in Section 4, we present the empirical results, where we apply the GWSF to the case of the US saving banks and compare it to a fixed-effects specification. Finally, we make our concluding remarks in Section 5.

2. Methodology

In this study, we employ two different specifications of the stochastic frontier model (SF). One model is the standard method that we estimate using fixed-effects. In the other model, we use a geographically weighted estimation process (GWE), in which we estimate distinct coefficients for each bank. Therefore, the GWE controls for the effects of the regional environment over the functioning of the firms (in this paper, banks). Our idea is to compare the results from these two specifications and, finally, to reach meaningful conclusions regarding the usefulness of the GWE in the efficiency estimation.

The basic SF model assumes that the production of a producer unit (company, government, machine, etc.) depends on the level of usage of the required inputs, of a normal random shock (and other uncontrollable factors) that affects the productivity of the unit and of other components associated with the inefficiency of the unit under managerial control. The latter always takes positive values, and we assume its distribution is strictly positive. The degree of efficiency represents how close a bank is in relation to the stochastic frontier.

Thus, we can describe the SF model, in its Cobb-Douglas version, as follows[5]:

$$AX_i = e^{\alpha_0} \left(\prod_{j=1}^{J} X_{ij}^{\alpha_j} \right) e^{\nu_i - \mu_i} \tag{1}$$

[5]For our purposes, the Cobb-Douglas version is more appropriate because it allows a more direct assessment of the elasticity of substitution for inputs and outputs as well as a clearer evaluation of the geographical intensity of these variables.

where AX_i is the output potentially generated by producer i; X_{ij} is the utilization level of input j by producer i; the α's are coefficients to be estimated; ν_i is the idiosyncratic term log-normally distributed with zero mean and standard deviation σ_ν, i.e., $\nu_i \overset{iid}{\sim} N(0, \sigma_\nu^2)$ and μ_i is the inefficiency component with distribution log-normal truncated in one and standard deviation σ_μ, i.e., $\mu_i \sim N^+(\overline{\mu}, \sigma_\mu^2)$.

Once we achieve this index, we can use it in the production of several distinct outputs and thus, in the case of the Cobb-Douglas function:

$$AY_i = e^{\beta_0} \prod_{k=1}^{K} Y_{ik}^{\beta_k}, \tag{2}$$

where AY_i denotes the amount of output effectively produced by producer i. From these two equations one can extract the efficiency level of each unit:

$$\theta_i = e^{\mu_i} = e^{\alpha_0 - \beta_0} \left(\prod_{j=1}^{J} X_{ij}^{\alpha_j} \right) \left(\prod_{k=1}^{K} Y_{ik}^{-\beta_k} \right) e^{\nu_i} \tag{3}$$

Here, θ_i is the efficiency level of unit i. As in the Cobb-Douglas case the homogeneity constraint over inputs holds if we normalize the variables in relation to one input. One can then write the following:

$$\begin{aligned} \frac{\theta_i}{X_{iJ}} &= e^{\gamma_0} X_{iJ}^{-1} \left(\prod_{j=1}^{J} X_{ij}^{\alpha_j} \right) \left(\prod_{k=1}^{K} Y_{ik}^{-\beta_k} \right) e^{\nu_i} \\ &= e^{\gamma_0} X_{iJ}^{-1+\Sigma_{m=1}^{J}\alpha_m} \left(\prod_{j=1}^{J} \left(\frac{X_{ij}}{X_{iJ}} \right)^{\alpha_j} \right) \left(\prod_{k=1}^{K} Y_{ik}^{-\beta_k} \right) e^{\nu_i} \end{aligned} \tag{4}$$

We can express this equation in its logarithmic version as follows:

$$\log \theta_i - (\Sigma_m \alpha_m)\, x_{iJ} = \gamma_0 + \sum_{j \neq J}^{J} \alpha_j x_{ij}^* - \sum_{k=1}^{K} \beta_k y_{ik} + \nu_i \tag{5}$$

where x and y are the neperian logarithms of X and Y, respectively, $x_j^* = \log \frac{X_j}{X_J}$ and $\gamma_0 = \alpha_0 - \beta_0$. We rearrange this equation to isolate input x_i:

$$-x_{iJ} = \gamma_0 + \sum_{j \neq J}^{J} \left(\frac{\alpha_j}{\Sigma_m \alpha_m} \right) x_{ij}^* - \sum_{k=1}^{K} \left(\frac{\beta_k}{\Sigma_m \alpha_m} \right) y_{ik} - \log \theta_i + \nu_i \tag{6}$$

One can write equation (6) in an estimable form as:

$$-x_{iJ} = \gamma_0 + \sum_{j \neq J}^{J} \left(\frac{\alpha_j}{\Sigma_m \alpha_m} \right) x_{ij}^{*} - \sum_{k=1}^{K} \left(\frac{\beta_k}{\Sigma_m \alpha_m} \right) y_{ik} + \nu_i - \mu_i \tag{7}$$

with: $\log \theta_i = \mu_i$.

We expect the α parameters to be positive because a reduction in the utilization of the reference input x_J must be compensated for by an increase in the utilization of the other inputs. Conversely, we expect the parameters β to be negative because a reduction in the utilization of the reference input will cause a decrease in the production, everything else being equal.

Taking this Cobb-Douglas function, the marginal rate of technical substitution between two inputs is given by the following:

$$MRTS_{i,j} = \frac{\alpha_j}{\alpha_i} \frac{X_i}{X_j} \tag{8}$$

Note that the marginal rate of technical substitution depends on the proportion in which the two inputs are being utilized. The ratio $\frac{\alpha_j}{\alpha_i}$ refers to the case in which both inputs are being utilized with the same intensity. This ratio can be recovered from the estimated values in equation (5) by dividing these figures by each other in pairs.

In the GWE, we apply the maximum likelihood method sequentially to each unit, and each separate observation gains a weight according to the geographical distance relationship to the reference unit. We assign these weights according to the following rule:

$$W_{ij} = \frac{e^{-\left(\frac{d_{ij}}{\lambda}\right)^2}}{\lambda \sqrt{2\Pi}}, \tag{9}$$

where W_{ij} is the weight of the j-unit in the estimation referenced over the i-unit; d_{it} is the great-circle distance in kilometers between the two units[6]; and λ is a dispersion parameter (bandwidth). If there are I units, we can then normalize the weights so that their sum becomes equal to I, as follows:

[6]Note that d_{it} does not need to be the geographical distance. For instance, it may be some sort of regulatory distance proxy, or even a linear combination between these two.

$$\varpi_{ij} = \frac{IW_{ij}}{\sum_{k=1}^{J} W_{ik}} \tag{10}$$

In each estimation, we multiply the normalized weights by their respective observations. Because we use all units as a reference in turn, we perform I estimations and estimate I sets of parameters, one for each unit.

The next step is to choose the appropriate λ. This parameter sets the weight distribution: the larger λ's magnitude, the greater the weight allocated the more distant units. The selection process is interactive, and we first establish a start value for it. In the algorithm that we create for this purpose, we use the standard deviation of the distances between the units as a point of departure. The algorithm then proceeds to estimate the geographically weighted stochastic frontier (GWSF) and to collect the mean sum of the squared residuals of the regressions that are obtained in the estimation process, which is the parameter to be minimized. We repeat the process with incremental variations in the bandwidth until the mean sum of the squared residuals ceases to decline, i.e., reaches it minimum.

When panel data is available, there are several different methods for estimating the inefficiency of the various producing units, see Kumbhakar and Lovell (2003) for a survey of these methods.

Because the fixed-effects model allows for correlation between the regressors and the inefficiency component and between the inefficiency component and the idiosyncratic shock, it appears to be the natural choice for us because the input distance is formed by a composition of these two stochastic terms $-\ln(D_{Ii}) = \mu_i - \nu_i$. In the panel data model, the equation to be estimated is as follows:

$$-x_{iJt} = \gamma_0 + \sum_{j \neq J}^{J} \left(\frac{\alpha_j}{\Sigma_m \alpha_m} \right) x^*_{itj} - \sum_{k=1}^{K} \left(\frac{\beta_k}{\Sigma_m \alpha_m} \right) y_{itk} + \nu_{it} - \mu_i \tag{11}$$

where the subscript t refers to time. This equation can be modified to the following:

$$-x_{iJt} = \gamma_{0i} + \sum_{j \neq J}^{J} \left(\frac{\alpha_j}{\Sigma_m \alpha_m} \right) x^*_{it,j} - \sum_{k=1}^{K} \left(\frac{\beta_k}{\Sigma_m \alpha_m} \right) y_{itk} + \nu_{it} \tag{12}$$

$$-x_{iJt} = \gamma_{0i} + \sum_{j \neq J}^{J} A_j x^*_{ijt} + \sum_{j=1}^{J} B_j y_{ijt} + \nu_{it} \tag{13}$$

with $\gamma_{0i} = \gamma_0 - \mu_i$, $A_j = \frac{\alpha_j}{\Sigma_m \alpha_m}$ and $B_j = \frac{\beta_k}{\Sigma_m \alpha_m}$. We then estimate this equation by OLS. To find γ_0, and consequently μ_i, we could use the following estimators:

$$\hat{\gamma}_0 = \max_i(\hat{\gamma}_{0i}) \tag{14}$$

$$\hat{\mu}_i = \hat{\gamma}_0 - \hat{\gamma}_{0i} \tag{15}$$

And we can obtain the technical efficiency of each unit with the following:

$$TE_i = e^{\hat{\mu}_i} \tag{16}$$

In the GWSF, we will have I sets of technical efficiencies, one for each weighted regression. In this case, the inefficiency of one particular producing unit will be the inefficiency we obtain in the regression that uses the specific unit as a reference for the weights calculation.

3. Database and variables

Our data is an unbalanced panel, which contains registers of 198 savings banks over nine years (2001-2009), totaling 1260 observations. We take the relevant data from BankScope, a financial database distributed by BVD-IBCA. Note that our data comprises two antagonic periods: one period of financial stability and bank consolidation until 2007 and another period of financial turmoil after 2008. This fact allows us to draw interesting conclusions on how efficiency evolves through the years. Our data comprises savings banks from over 43 states and 172 towns. Map 1 indicates the dispersion of these banks across the US.

[Map 1]

Because the panel is unbalanced, there are banks that are not present in every year of the analysis due to the beginning/ending of operations or to missing data. Table 1 shows the number of banks present in the data set each year.

[Table 1]

We choose the following variables for this exercise: personnel expenses (input), interest expenses (input), other expenses (input, defined as total expenses minus personnel expenses minus interest expenses), bank loans (output), liquid assets (output), total deposits (output), and non-interest income (output). Personnel expenses and interest expenses are the usual measures of bank costs and are commonly employed as input variables. Other expenses, as stated above, are the remaining expenses taken altogether. This division into categories of expenses occurs because there might be an optimal combination of expenses that enhances productivity. Most of the output variables are also quite traditional because the literature vastly utilizes bank's loans, liquid assets and total deposits as outputs. The inclusion of non-interest income aims to capture the non-traditional bank activities, which are supposed to be quite distinct in geographical terms[7]. Even recognizing that the other variables could be geographically dependent, we believe that the introduction of the non-interest income might be more appropriate to capture this aspect.

[Table 2]

Table 2 presents some descriptive statistics of the variables. Loans and deposits are the main outputs of US saving banks, as this table shows. The most important input is interest expenses, because the banks' main activity is to intermediate interest-bearing funds. However, personnel expenses account for approximately 18% of the expenses on average. This proportion is high and it demonstrates that saving banks are more labor intensive than regular commercial banks. Finally, one can observe that all of the variables have a high standard deviation, which indicates that we consider very heterogeneous banks in our specifications.

[7]Even though there is not a consensus on the matter yet, the literature attributes a great deal of importance in incorporating variables representing non-traditional bank activities (such as off-balance sheet and non-interest income) in the analysis of bank efficiency (Lozano-Vivas and Pasiouras, 2010). Ignoring these measures can be misleading because the analysis then does not take into account the bank's balance sheet as a whole

4. Empirical results

This section presents our model's empirical results. First, as previously stated in section 2, we estimate an input-distance function using the banks' other expenses as a reference (endogenous) variable. We perform this estimation in a non-spatial context using a fixed-effect OLS. Our choice for the reference input was based on the fact that the banks' other expenses is a variable that could include several distinct components and thus, the analysis of its coefficient would not be very informative.

We also include a time trend and its square term in the estimated equation to capture nonlinearities in the temporal tendency of the banks' efficiency. Thus, the regression can be written as follows:

$$-x_{it,1} = \alpha_1 + \alpha_2 x^*_{it,2} + \alpha_3 x^*_{it,3} + \beta_1 y_{it,1} + \beta_2 y_{it,2} + \beta_3 y_{it,3} + \beta_4 y_{it,4} + \beta_t t + \beta_T t^2 + \nu_{it} - \mu_i \quad (17)$$

where $x_{it,1}$ is the logarithm of the banks' other expenses; $x^*_{it,2}$ is the logarithm of the ratio between the personnel expenses and banks' other expenses; $x^*_{it,3}$ is the logarithm of the ratio between the interest expenses and the banks' other expenses; $y_{it,1}$ is the logarithm of the bank's loans; $y_{it,2}$ is the logarithm of the total liquid assets, $y_{it,3}$ is the logarithm of the total deposits and $y_{it,4}$ is the logarithm of the non-interest income. The use of a non-traditional activity output is necessary to avoid bias in the results. According to Lozano-Vivas and Pasiouras (2010), these activities have increasing importance in the bank's balance sheet. t and t^2 are, respectively, the year of the observation and its square. Table 3 shows the results of the estimated equation.

[Table 3]

The values of the R^2 and the F-statistic are satisfactory, and indicate a good fit for the model. The t-statistics suggest the significance of the parameters for all variables except for total liquid assets and the time trend. We will next provide a further analysis of the interpretation of this model's coefficients. The purpose of this analysis is to determine how each variable affects the employment of the reference input. We then perform a comparison of the fixed effects estimator against the geographically weighted model.

It is clear that all input and output coefficients possess the expected signs. The input variables' positive values mean that, the greater utilization of any input, the smaller the necessity of utilizing the reference input[8]. In other words, given a determined amount of bank input, the use of one additional unit of input X_1 means the lower employment of inputs X_i, $\forall\ i \neq 1$ in one unit. Reciprocally, the greater the production of any output, the greater the utilization of the reference input (everything else being constant). It is reasonable to suppose that the production of one additional unit of output requires more inputs in general.

Another inference from the values of the coefficients is that US saving banks appear to have increasing returns to scale. The scale elasticity is equal to the ratio between the sum of the coefficients of the input variables, plus one, and the sum of the absolute values of the coefficients of the output variables. In this case, the estimated scale elasticity for the non-spatial model was 2.18. This result implies that US saving banks have not yet achieved their optimal size in terms of technical efficiency.

The coefficient associated with the time trend is positive, but not significant, while the coefficient linked to its square term is negative and significant. This result can mean either that there is a negative evolution in the banks' efficiency or that the coefficient of the time trend is truly positive, but the estimation fails to recognize this fact. Fortunately, the results of the GWE suggest that the latter possibility is probably true.

Additionally, a traditional Translog transformation function was estimated to test for model robustness. The correlation between the Cobb-Douglas and the Translog results was 0.79, which indicates a good adherence between the two models results. The Cobb-Douglas estimation indicates an average efficiency of 0.462, while the Translog model produces an efficiency mean of 0.350.

Our second approach is the GWE. This process involves 198 sub-estimations, each of them with a distinct set of weights. We obtain the change in the weights by varying the standard deviation of the normal generating function. We began with a standard deviation of 100 km and went up to 15,000 km. The better parameter, determined by the minimum average of the square residuals sum, is equal to 2,400km. In graph 1, we present the relationship

[8]Remember that the reference variable is taken in negative values.

between the standard deviation of the weight-generating function (λ) and the exponential of the average sum of the square residuals.

Once more a Translog model was tested, this time using geographical weights. The comparison with the results of the Cobb-Douglas estimation also points to the model's robustness. The correlation between the banks technical efficiency as we obtain for the two methods was 0.85, and the correlation between the efficiency ranks was 0.9.

[Graph 1]

Table 4 pictures the results of the GW estimation. This table presents some descriptive statistics of the coefficients that we estimate in all of the local sub-estimations. To clarify any possible confusion, the standard deviations in the third row are not the parameter associated with the estimator but are the deviations of the estimated values. The last row once again shows the results of the non-spatial estimation to facilitate the comparison. One remarkable fact is that all of the coefficients are significant in the GW estimation.

[Table 4]

All of the estimations also provide evidence of increasing returns to scale, although somewhat smaller than those we obtain in the non-spatial estimation (1.98 against 2.18). This result means that the fixed effects model might overestimate (in relation to the GWE) the aggregate effects of the inputs on the reference and/or might underestimate the impact of the output on the bank's other expenses. The optimal size of US saving banks is somewhat lower when we control for geographical factors.

The GWE clarifies the time trend for efficiency, which is increasing but with diminishing returns. That is to say that in the last years of the sample, the trend's inclination decreases. This finding is consistent with the impact of the financial crisis on US savings banks, particularly small ones.

Comparing the estimated coefficient of the spatial and non-spatial frontiers, some interesting insights can be drawn. First, the coefficient associated with personnel expenses (α_2) is always greater in the non-spatial estimation in relation to the GWE. Map 2 shows that α_2 varies from 0.453 to 0.461, while in the non-spatial case it equals 0.545. This difference implies a greater marginal rate of technical substitution between the personnel expenses and the other expenses in the former than in the latter.

[Map 2]

The inverse appears to occur with the interest expenses. As one can observe in Map 3, their GW coefficients always have greater values than those for the non-spatial case.

[Map 3]

The significance of these results is that once you control for geographical factors, it is easier to substitute for manpower with other types of inputs, but it is harder to do the same for loaned funds.

In terms of outputs, the major distinction that emerged regards the total deposits. This item appears to be definitively more relevant in the spatial case. The larger coefficient associated with this variable implies that a larger number of deposits is necessary to generate a given amount of other outputs at the regional level. This effect points out the importance of local branches in the capture of deposits.

[Map 4]

The non-interest income, contrary to expected, did not vary significantly between the spatial and non-spatial models. Indeed, the non-spatial case showed a slightly greater coefficient, although the difference between them is not significant.

[Map 5]

When the technical efficiencies from the two models are contrasted, one can observe that 126 banks had their efficiency improved in the spatial model. This phenomenon occurs because the model compares banks with others near by that are probably subject to similar opportunities and constraints. One direct implication is that the comparisons are more flexible in this model. That is, a bank's efficiency is not measured against a standard established by faraway institutions of very distinct nature and subject to different environments. Therefore, a bank that is not considered to be efficient in comparison to the best practice bank in the sample might be very efficient in relation to others in its vicinity.

In spite of this large proportion of banks whose efficiency the spatial model improves, there were no substantial changes in the efficiency average or the standard deviation from one model to the other. This result means that, once we define a spatial model, efficiency was, to a large extent, redistributed from those who presented a higher performance in the FE model to those with lower scores.

[Map 6]

[Map 7]

A comparison between maps 6 and 7 permits an analysis of the differences in efficiency in the spatial and non-spatial models. It is clear that the banks from the Northeast US appear to be more efficient in the GWSF model. To the contrary, the west and the south of the country have banks with a lower efficiency in the spatial case.

4.1. Robustness test

In order to check if our technical efficiency model is consistent with other popular measures of estimating efficiency, we also estimate a geographically weighted cost function. According to Bauer et al. (1998), efficiency scores should: (i) have comparable summary statistics; (ii) rank institutions similarly; (iii) point out the same best and worst practice banks; (iv) be consistent over time; and (v) consistent with nonfrontier efficiency methods. It is worthwhile to note that, it is not possible to check these conditions exactly as proposed by Bauer et al. (1998), since that out of the four suggested approaches (DEA, SFA, TFA (Thick frontier analysis) and DFA (Distribution-free approach)), the Stochastic Frontier is the only one with a geographically weighted version.

As we deemed that comparing a geographically weighted model with traditional ones would be inappropriate, we tried a adapted version of Bauer et al. (1998), in which we use the comparing parameters (efficiency scores, efficiency rankings and so on) of three GW estimations (Cob-Douglas, translog and cost function).

At the specific case of the GW cost function, input variables were aggregated to compose the total cost of the bank. It was possible to do so, since al input variables were denominated in current dollars. We estimate cost inefficiency similarly as the technical inefficiency case in equation (11). However, the composed error term is $\nu_{it} + \mu_i$ and the left hand side variable is total costs.

Table 5 shows the consistency tests that we perform. In this table, one can notice that GW methods preserve the bulk of results across specifications. The efficiency ranks correlations with the benchmark model are equal to 89.4% for the translog GW model and 75.6% for the cost function model. In addition, 91% of the banks deemed efficient at the Cobb-Douglas GW model also showed unitary score in the translog GW model and 84% in the GW cost function model.

[Table 5]

5. Conclusions

This paper applies a geographically weighted stochastic frontier model to a panel data set of US saving banks to determine their efficiency levels between 2001 and 2009. Because there is a significant evidence that environment influences the estimation of efficiency frontiers, we control the environment through the GWSF method. This method has never been employed in panel data until now, particularly in the banking literature. As evidence of the advantages and viability of the GWSF, we compare the results of this method with a fixed-effects SF model.

Many studies apply the regular SF model and control for unobservable factors using macroeconomic variables or even a fixed effects model. The problem of the former is related to the choice of variables to use in the specification. In other words, one might not be certain about which macroeconomic factors have a significant effect on bank efficiency and there could be several unobservable variables that do not have a suitable proxy. In addition, this approach can fail to control for the local characteristics. We show that the latter, however, by eliminating the specification time-invariant factors, can sub-estimate efficiency by itself. In addition, the environmental characteristics that have changed in a particular period of time would not be captured by the fixed effect approach.

The case of US saving banks, the subject of our analysis, is evidence to support the GWSF approach. These banks operate on a regional level by lending to small and medium enterprises and thus, local characteristics can have a higher influence in their behavior than a bank that operates across the entire country. When we control for geographical factors, the technical efficiency appears to be larger for most of the banks because the estimation gives the banks that operate closer to a specific bank a higher weight. In other words, the bank performance is now compared to those banks that are subject to the same constraints and not to banks that have completely different conditions. Our overall conclusion is that geography matters, and it plays an essential role in correctly estimating efficiency. This result has important implications for policy makers because one policy does not necessarily fit all.

We highlight some of the secondary findings next. (i) We find a positive time trend of efficiency for the period in the GWSF model. However, this positive trend decreases as the period under consideration nears its end. This decrease is evidence of the effects of financial turmoil on US bank efficiency that the FE model fails to capture. (ii) There are differences between the models in terms of both the inputs and the outputs' estimated coefficients, as well, with the exception of non-interest income. The substitution between this last variable and the reference input appears to be insensitive to the local conditions and constraints.

Further analysis could use similar weighing methods to estimate cross-country bank efficiency with the purpose of comparing the results from this approach with the results from the methods that have been used so far, i.e., the employment of country-specific environmental variables. Note, however, that in international comparisons it might not be enough to consider only geographical distance in the weights estimation, but other measures of interconnectedness must also be considered.

References

Aigner, D., Lovell, C. A., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6 (1), 21–37.

Akhigbe, A., McNulty, J. E., 2003. The profit efficiency of small US commercial banks. Journal of Banking & Finance 27, 307–325.

Battese, G. E., Coelli, T. J., 1992. Frontier production functions, technical efficiency and panel data with application to paddy farmers in India. Journal of Productivity Analysis 3, 153–169.

Battese, G. E., Coelli, T. J., 1995. A model for technical inefficiency effects in a stochastic frontier production function models. Empirical Economics 20, 325–332.

Bauer, P. W., Berger, A. N., Ferrier, G. D., Humphrey, D. B., 1998. Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. Journal of Economics and Business 50, 85–114.

Berger, A. N., 2003. The economic effects of technological progress: Evidence from the banking industry. Journal of Money, Credit and Banking 35, 141–176.

Berger, A. N., DeYoung, R., 2001. The effects of geographic expansion on bank efficiency. Journal of Financial Services Research 19, 163–184.

Berger, A. N., Hasan, I., Zhou, M., 2009. Bank ownership and efficiency in China: What will happen in the world's largest nation? Journal of Banking & Finance 33, 113–130.

Berger, A. N., Humphrey, D. B., 1997. Efficiency of financial institutions: International survey and directions for future research. European Journal of Operational Reserch 98, 175–212.

Berger, A. N., Leusner, J. H., Mingo, J. J., 1997. The efficiency of bank branches. Journal of Monetary Economics 40 (1), 141–162.

Bos, J. W. B., Kool, C. J. M., 2006. Bank efficiency: The role of bank strategy and local market conditions. Journal of Banking & Finance 30, 1953–1974.

Bos, J. W. B., Schmiedel, H., 2007. Is there a single frontier in a single European banking market? Journal of Banking & Finance 31, 2081–2102.

Degryse, H., Ongena, S., 2004. The impact of technology and regulation on the geographical scope of banking. Oxford Review of Economic Policy 20 (4), 571–590.

DeYoung, R., Klier, T., McMillen, D. P., 2004. The changing geography of the U.S. banking industry. The Industrial Geographer 2, 29–48.

Dietsch, M., Lozano-Vivas, A., 2000. How the environment determines banking effciency: A comparison between French and Spanish industries. Journal of Banking & Finance 24, 985–1004.

Hasan, I., Koetter, M., Wedow, M., 2009. Regional growth and finance in Europe: Is there a quality effect of bank efficiency? Journal of Banking & Finance 33, 1446–1453.

Kumbhakar, S. C., Lovell, C. A. K., 2003. Stochastic Frontier Analysis. Cambridge Books. Cambridge University Press.

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

Lozano-Vivas, A., Pasiouras, F., 2010. The impact of non-traditional activities on the estimation of bank efficiency: International evidence. Journal of Banking & Finance 34, 1436–1449.

Lozano-Vivas, A., Pastor, J., Pastor, J. M., 2002. An efficiency comparison of European banking systems operating under different environmental conditions. Journal of Productivity Analysis 18, 59–77.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444.

Paradi, J. C., Rouatt, S., Zhu, H., 2011. Two-stage evaluation of bank branch efficiency using data envelopment analysis. Omega 39 (1), 99–109.

Paradi, J. C., Schaffnit, C., 2004. Commercial branch performance evaluation and results communication in a Canadian bank - a DEA application. European Journal of Operational Research 156, 719–735.

Pasiouras, F., Sifodaskalakis, E., Zopounidis, C., 2011. The cost efficiency of Greek cooperative banks: An application of two-stage data envelopment analysis. International Journal of Financial Services Management 5 (1), 3451.

Portela, M. C. A. S., Thanassoulis, E., 2007. Comparative efficiency analysis of portuguese bank branches. European Journal of Operational Research 177, 1275–1288.

Samaha, S. A., Kamakura, W. A., 2008. Assessing the market value of real state property with a geographically weighted stochastic frontier model. Real Estate Economics 36 (4), 717–751.

Tirtiroglu, D., Tanyeri, A. B., Tirtiroglu, E., Daniels, M. K. N., Apr. 2011. Banking geography and cross-fertilization in the productivity growth of us commercial banks. Working Papers 1108, TUSIAD-Koc University Economic Research Forum.

Table 1: Number of Banks in the Data set by Year

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Banks	81	198	198	170	162	152	135	122	109

Table 2: Descriptive Statistics of the Utilized Variables (in US$ thousands)

Variables	Status	Mean	Median	St. Dev.	Max.	Min.
Personnel Expenses	Input	45.142	16.388	88.932	810.737	809
Interest Expenses	Input	120.028	31.050	289.004	4.710.007	342
Other Expenses	Input	90.183	15.594	356.884	6.087.496	555
Bank's Loans	Output	3.607.313	865.508	9.487.042	125.167.453	23.428
Liquid Assets	Output	158.349	45.010	455.762	7.068.700	1.164
Total Deposits	Output	2.895.914	920.398	6.117.170	69.603.422	41.658
Non-interest Income	Output	72.070	9.616	270.300	4.098.312	63

Source: BankScope.

Table 3: Fixed-Effect, Non-spatial, OLS Estimation

Variable	$\mathbf{x}_2^*$	$\mathbf{x}_3^*$	$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$	$\mathbf{y}_4$	$\mathbf{t}$	$\mathbf{t}^2$
Coefficient	0.545	0.237	-0.351	-0.011	-0.296	-0.159	0.008	-0.005
St Dev	0.021	0.017	0.033	0.009	0.035	0.011	0.01	0.001
t-stat	25.902	14.032	-10.793	-1.196	-8.331	-14.479	0.778	-4.579
p-value	0	0	0	0.232	0	0	0.437	0
$R^2 = 0.872$			F -stat = 899.762				F p-value = 0.000	

Table 4: Descriptive Statistics of the Estimated GWE Coefficients

Coefficient	α_2	α_3	β_1	β_2	β_3	β_4	β_t	β_T
Mean	0.459	0.282	-0.345	-0.027	-0.359	-0.144	0.02	-0.006
Median	0.46	0.281	-0.345	-0.027	-0.359	-0.145	0.02	-0.006
Stand. Dev	0.002	0.001	0.001	0.001	0.001	0.001	0	0
Max	0.461	0.285	-0.344	-0.027	-0.358	-0.142	0.021	-0.006
Min	0.453	0.281	-0.348	-0.029	-0.36	-0.145	0.02	-0.006
Non-spatial	0.545	0.237	-0.351	-0.011	-0.296	-0.159	0.008	-0.005

Note: this table presents the summary statistics from the set of local regressions in the GWE approach. This method basically consists in the estimation of a local frontier for each bank, while the other banks are weighted by their distance in relation to the reference bank. Inefficiency of one particular producing unit will be that obtained in the regression in which this specific unit was used as reference for the weights calculation. We present the coefficients of Table 3 (fixed effects) as a mean of comparison.

Table 5: Consistency tests based on Bauer et al. (1998)

Method	Efficiency Score		Rank correlation with Cobb-Douglas GW	Proportion of coincident banks on the frontier in relation to the Cobb-Douglas GW model
	Mean	SD		
Cobb-Douglas GW	0.574	0.237	-	-
Translog GW	0.582	0.189	0.894	0.91
Cost Function GW	0.498	0.325	0.756	0.84
Non-spatial Cobb-Douglas	0.462	0.209	0.843	0.87

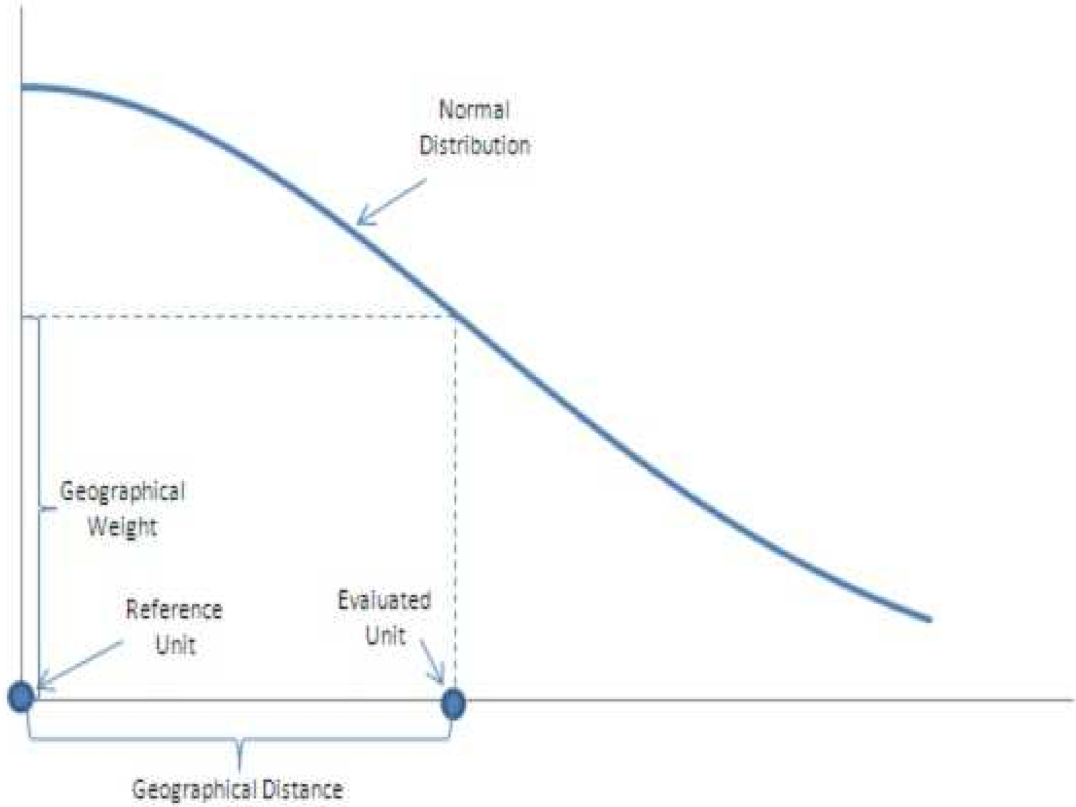


Figure 1: *Graph1*

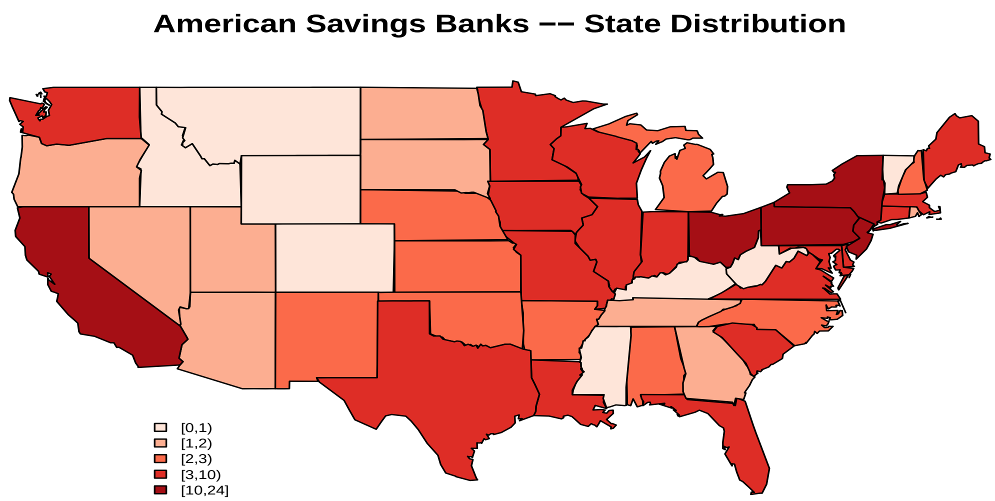


Figure 2: *Map1*

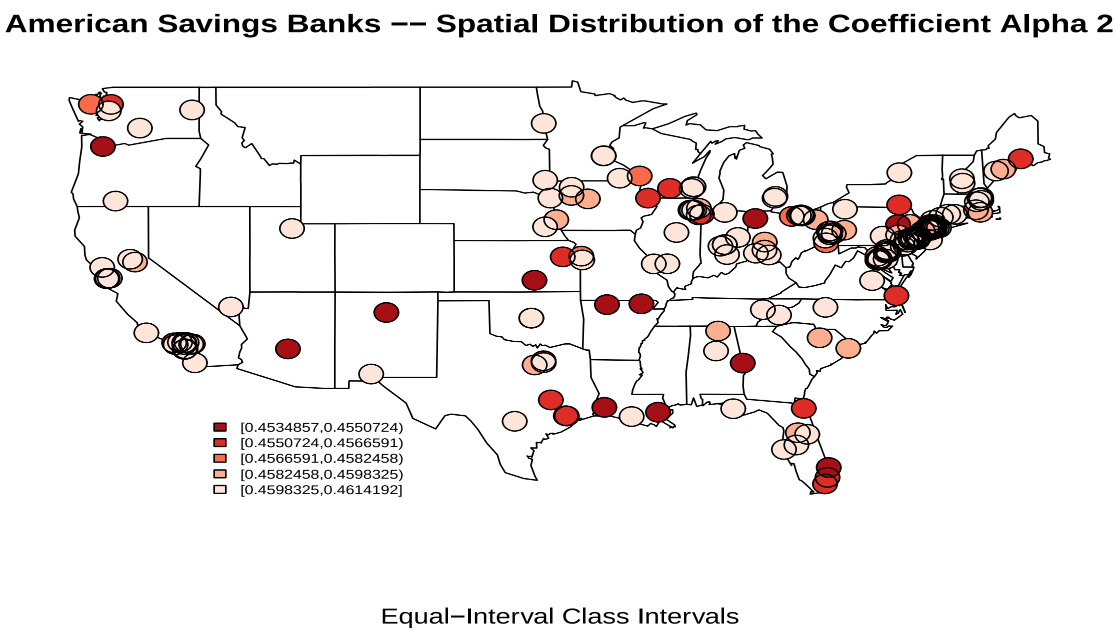


Figure 3: *Map2*

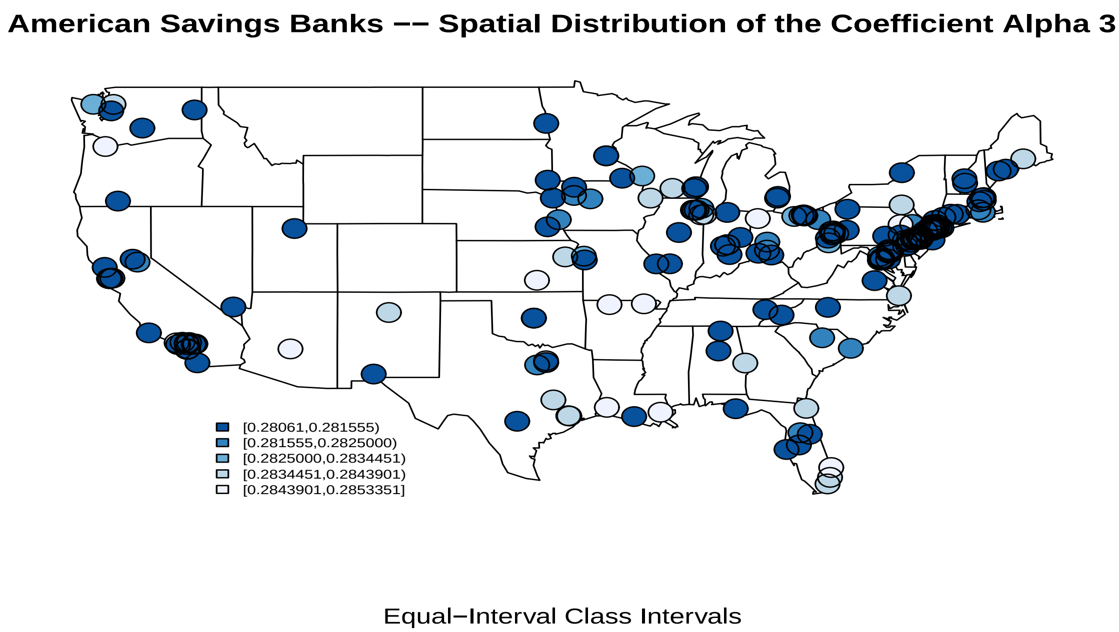


Figure 4: *Map3*

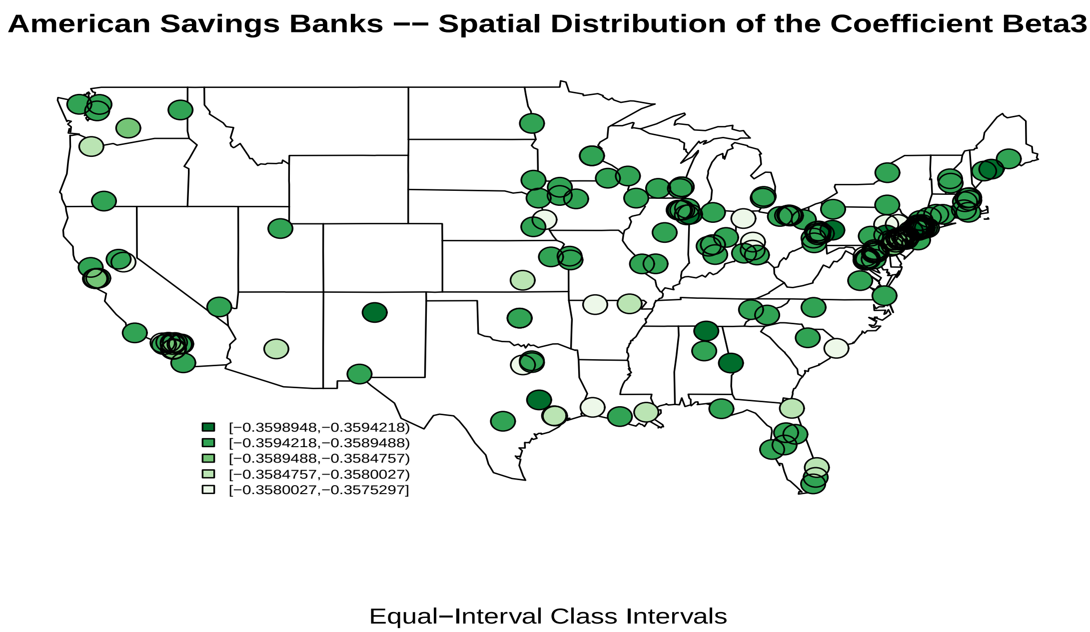


Figure 5: *Map4*

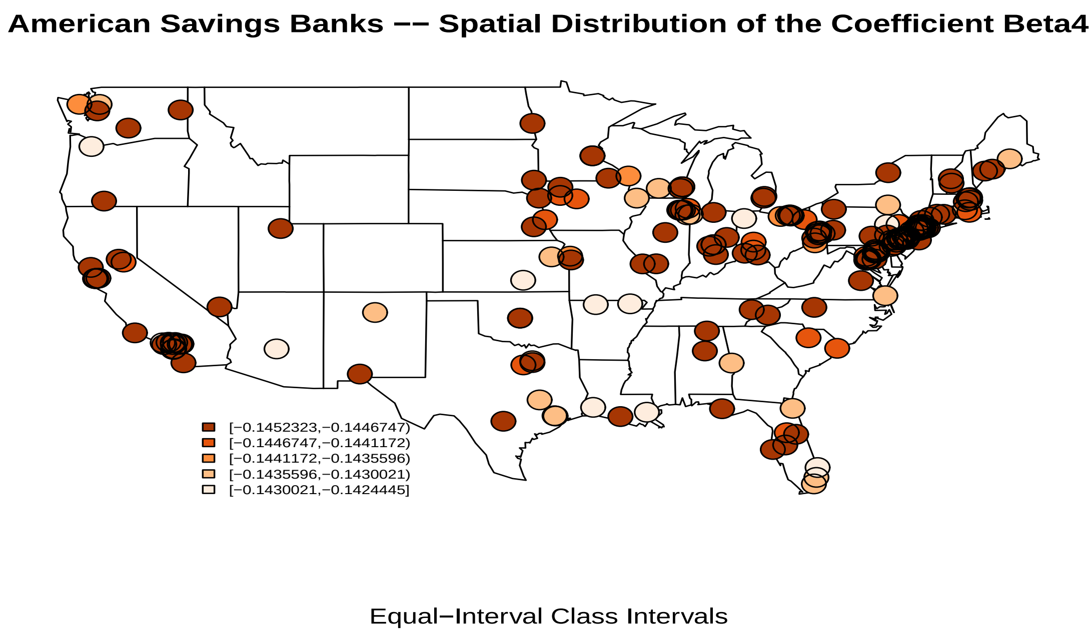
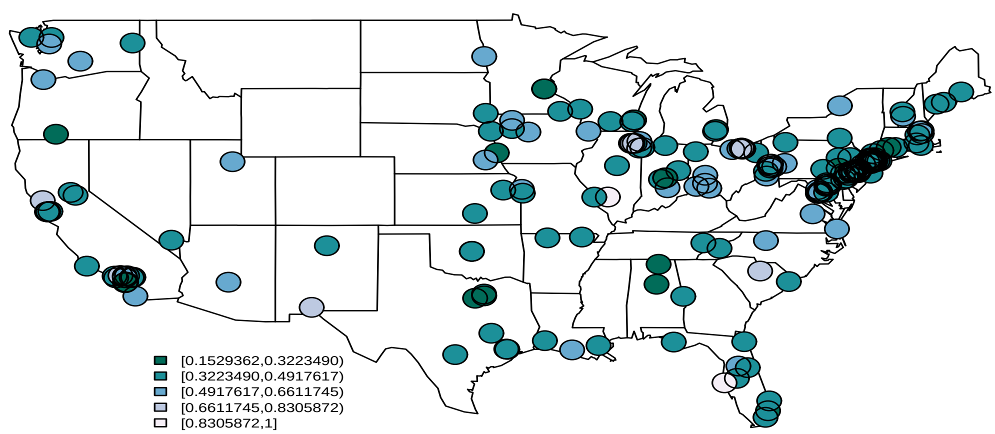


Figure 6: *Map5*

American Savings Banks –– Technical Efficiency in the Non–Spatial Model



Equal–Interval Class Intervals

Figure 7: *Map6*

American Savings Banks –– Technical Efficiency in the Spatial Model

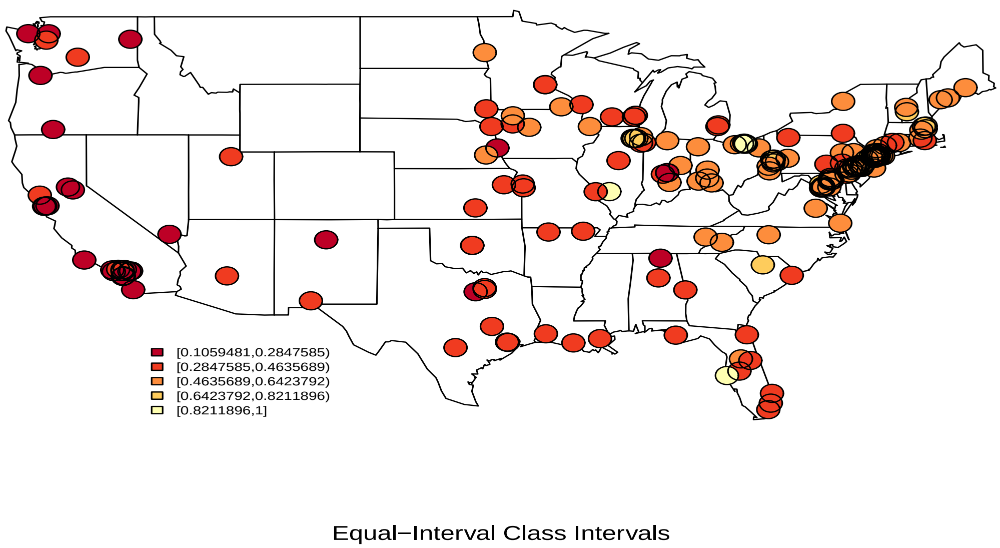


Figure 8: *Map7*