MEASURING THE INFLUENCE OF AGGLOMERATION EFFECTS ON THE U.S. BANKING INDUSTRY EFFICIENCY

by

Kanel Polina

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis
.

Kyiv School of Economics

2015

Thesis Supervisor: _____ Professor Vakhitov Volodymyr

Approved by
Head of the KSE Defense Committee, Professor Irwin Collier

Date _____

Kyiv School of Economics

Abstract

MEASURING THE INFLUENCE OF AGGLOMERATION EFFECTS ON THE U.S. BANKING INDUSTRY EFFICIENCY

by Kanel Polina

Thesis Supervisor: Professor Vakhitov Volodymyr

Decision regarding branch location became strategic, which should accounts for the unique surrounding of the given location and the dynamic of the external economic factors. This paper aims to estimate the spillover effect from the branch surrounding due to the agglomeration of the branches in the New York Metropolitan statistical area. After aggregating data from the branch to the 5-digit zip code level data, it was found that indeed there is spillover effect with elasticity around 0.13, which is realized in the additional efficiency gain in the deposit collection process. Also it was found that the services (non-finance) industry development has significant influence on the region branch efficiency with elasticity of 3.5. At the same time the diversity of the branch or the service provided types in the region do not have significant influence on the deposits collection efficiency.

TABLE OF CONTENTS

LIST OF FIGURES ii

LIST OF TABLES iii

ACKNOWLEDGEMENTS iv

GLOSSARY v

Chapter 1: INTRODUCTION 1

Chapter 2: LITERATURE REVIEW 4

2.1 Agglomeration drivers and effects in the financial industry 5

2.2 The market border in the agglomerations 7

2.3 The efficiency measure 9

Chapter 3: DATA DESCRIPTION 12

3.1 Description of the industry 12

3.2 Control variables 13

Chapter 4: METHODOLOGY 17

Chapter 5: RESULTS 22

Chapter 6: CONCLUSIONS 26

WORKS CITED 38

LIST OF FIGURES

Number *Page*

Figure 1. The development of the branching in the U.S. from 1984-2003……29

Figure 2. The distribution of the branch network in the U.S……29

Figure 3. The level and the share of the real financial GDP in the total real GDP for the ten largest MSA……..30

Figure 4. The number of branches per 5-digit zip code in the New York MSA…….....30

LIST OF TABLES

Number *Page*

Table 1. The aggregated bank branches data in New York MSA 31

Table 2. Definition of the services and branch types 32

Table 3. The 5-digit zip code level economic data for the New York MSA.....33

Table 4. The description of the economic activity density in the region 33

Table 5. Moran I statistics for New York MSA 33

Table 6. Regressions estimation for the New York MSA 34

Table 7. Model specification test results 35

Table 8. Elasticity of the deposits density with respect to the explanatory variables at mean.....35

Table 9. Regressions estimation with service type dummies 36

Table 10. Model specification test results with services dummies 37

Table 11. Elasticity with service type dummies 37

ACKNOWLEDGMENTS

The author wishes to express gratitude to the thesis adviser, Prof. Volodymyr Vakhitov for the valuable insights and suggestions that helped during the research; to Prof. Tom Coupe for the valuable comments and to my family and friends for their support and patience.

GLOSSARY

Agglomeration (conurbation)- a region comprising a number of cities, large towns, and other urban areas that, through population growth and physical expansion, have merged to form one continuous urban and industrially developed area.

Economies of agglomeration - the benefits that firms obtain by locating near each other ('agglomerating').

Deposits (bank deposit; deposit account) - money placed into a banking institution for safekeeping; bank deposits are made to deposit accounts at a banking institution, such as savings accounts, checking accounts and money market accounts.

Bank branch – a retail location where a bank, credit union, or other financial institution (and by extension, brokerage firms) offers a wide array of face-to-face and automated services to its customers.

Metropolitan statistical area (MSA) – a geographical region with a relatively high population density at its core and close economic ties throughout the area. Such regions are neither legally incorporated as a city or town would be, nor are they legal administrative divisions like counties and states.

Financial Real GDP – a macroeconomic measure of the value of financial industry output adjusted for price changes (i.e., inflation or deflation).

Chapter 1

INTRODUCTION

The modern financial services industry operates differently from the most of other economy sectors. Among distinctive features of the financial industry are dependence on local resources, such as human capital and economic development of the region. The availability of human capital is one of the crucial factors, since it directly influences the revenues and quality of the banking operations; the economic development performs the leading role in the banks industry development, since banks require the investment activity, flows of capital, as well as landing and borrowing operations. In addition, the availability of new technologies and instant access to information are becoming more and more important in the everyday operations nowadays. On the other hand, financial sector of the U.S. appeared to play an important role in the economic development. Roussea and Sylla (1999) found that power of modern growth of the U.S. economy came from the financial sector changes.

The banking industry tends to agglomerate, to create clusters of the activity within the certain region, in order to receive benefits from developed technology infrastructure, shared pool of skilled labor and information. These tendencies may well lead to the explicit realization of the spatial agglomeration effects, which are much more evident in financial industry in comparison with other.

The U.S. banking industry is one of the most developed in the world with the developed network of branches. From 1984 to 2003 the banking industry decreased twofold in terms of the number of banks, however, the total banks'

assets has increased threefold – from USD 307 mln in 1984 to USD 979 mln in 2003 (Garrett at el. 2003), in addition the number of branches is continue to grow significantly, what can be seen on the Figure 1. Among numerous opportunities, these tendencies also raise many concerns regarding the performance and efficiency of branching. New York is of special interest - while it follows the common trends, it is also the most powerful in terms of financial real GDP among other regions (see Figure 3).

Concerns about the efficiency are not without grounds. The inefficiency may lead to detrimental consequences for the whole economy. The overcrowded local market may lead to the inefficient allocation of the capital goods and investments, as well as the undesirable effects for the local economy. This in turn may well lead to the lower margins due to the tough competition, and lower profit level. On the other hand, the local markets with only several banks presented will lead to the oligopoly structures, when banks will be able to earn markups over the industry average, what finally will lead to the decrease in the social welfare (Richards at el. 2007). The definition of the market, the market structure and the competition has significant influence on the bank and branch performance. Therefore branching is no longer a random choice, but a strategic decision, which requires in-depth analysis of the branch efficiency behavior depending on the surrounding factors.

In the paper the ability of the banks to collect deposits is taken as an efficiency measure. Under the ability to collect the deposits the density per square feet is considered. The deposits as the measure of the efficiency is widely criticized, however other approaches are lack of the consistent data and are often do not depict the true revenues or require a lot of assumptions. Therefore, in this case

deposits density is considered an invariant measure of efficiency within the branches in the region.

There has been little empirical research on bank branch efficiency in the U.S. agglomerations. Previous research does not consider any spatial relationships between the branches. The independence of branch performance from the competitors one was omitted in the papers, what is the strong assumption regarding the branches competition. This paper tries to relax the assumption about the spatial indifference in branch efficiency and to determine how the performance of the branch may be affected by the surrounding factors.

The primary hypothesis of the paper is that there is a convincing evidence for agglomeration gains in the banks' efficiency to create the deposit inflow. These gains are created due to the shared pool of skilled labor, the development of the services industries, which support banking operations, and economic development of the region. Moreover, it is suggested that the structure of the market and the diversity of the services do not significantly influence on the level of the spillover gains.

The remainder of the paper is organized as follows. Chapter 2 analyzes the relevant literature about the agglomeration effects in the banking industry and spatial considerations. Chapter 3 describes the sources of the data and controlling variables. Chapter 4 introduces the methodology used for determining and measuring of the spatial relations. Chapter 5 discusses the obtained results and Chapter 6 summarizes findings and suggests the directions of the further research.

Chapter 2

LITERATURE REVIEW

Research on agglomeration economies has started from the studies of Alfred Marshall (1920). The agglomeration refers to the high concentration of the business entities in the certain geographic location, e.g. cluster or business centers. He considers three types of externalities that are essential in grouping industrial activities: shared labor pool, technology diffusion and development of the information exchange between entities, existence of the specialized services and inputs markets.

Agglomeration economies are understood as those gains or advantages obtained by companies, individuals or consumers due to their physical proximity to other firms, workers and consumers, proximity which is not available, at least to some degree, to other locations. The general idea behind this concept is that the operating environment has a positive impact on firm productivity.

Hoover (1948), Krugman (1991) and Malmberg (2009) defined two types of the agglomeration economies:

1) localization economies, which describe these advantages as a result of the spatial grouping of similar or related firms under the shape of industrial clusters; and
2) urbanization economies, which refer to the benefits obtained from the localization of a firm in a large and dense urban area.

2.1 Agglomeration drivers and effects in the financial industry

The evidences of agglomeration forces in the financial industry are found in many papers. Kindleberger (1973) and Gehrig (1998) study the phenomenon of the "financial center" and found that it is a typical structure for this kind of industry. The evidences of three agglomeration forces proposed by Marshall are found in finance industry by Krugman (1991). He states that such agglomeration effects as knowledge spillover and localization are much more apparent in the financial industry. Hall and Appleyard (2009) also find that highly skilled financiers (shared pool of skilled labor) and business knowledge are important factors in the knowledge spillovers in London's financial center. Moreover, according to Keeble and Nachum (2002), knowledge-intensive industries such as the financial sector should aim to benefit from agglomeration effects, which are created owing to the knowledge accumulation and innovation environment.

According to Birkin et al., (2002) strategic planning and decision making in financial industry resembles that of in other industries. In fact, the location of the branches is a significant part of the bank strategy, which requires to consider all drivers of the branch growth, such as spatial distribution of clients, distribution of own and competitor's branches, as well as the distribution of the inputs such as labor force and technology development of the target region. These distributions are found to be not homogeneous in space (Chang et al. 1997), what indicates that the position of the bank branch can be either strategic advantage or disadvantage.

Besides some apparent advantages such as unsaturated markets and rich clients, there are also agglomeration effects associated with the costs side. For instance, the agglomeration for bank branches provide some advantages such as reducing

bank customer acquisition costs. For consumers it is cheaper to find and to obtain banking services, on the other hand banks can share services and maintenance costs with other branches. Moreover, there is advantage from experience sharing, such as know-how and decision sharing, lower costs of the strategic market share maximizing (such as high margin products) due to the development of the market by competitors, etc.

One of the features of the banks spatial strategy is the proximity to the customers (Kutler, 1996). The distance to the bank plays one of the most important roles, as people tend to choose bank in the neighborhood for the deposits and cash operations. Thus, taking into account customers' preferences, bank cannot be located in any place or be substituted with the ATM (Chang et al. 1997). Therefore, the density of the population in the chosen location plays an important role in the deposits efficiency collection.

Other three key factors bank is looking for in the suggested locations are: access to the target clients, the potential for the increase in banking sales and the total banking earnings. It is worth noticing, that the first factor is underlying for the second and the third ones, since clients generate the cash flows. Therefore, such spatial factors as employment and the income of the people in the region play one of the most crucial roles. Employment concentration varies dramatically across space (Desmet and Rossi-Hansberg 2009), what is definitely affecting the spatial distribution of the branches. Moreover, income of the people is linked to their location (World Bank 2009), what indicate on the close relationship between the economic development, banks' opportunities to rise revenues and the income of the employees.

In analysis of spatial relationships, the densities of the inputs more important than the levels, since density can give the general picture of the market saturation and give the estimation of how intense relationships within the market may be. The works of Ciccone and Hall (1996) and Harris and Ioannides (2000) attempt to test the impact of employment density on productivity. The first uses data regarding economy in general, and the second, regarding the American metropolis; both found a positive effect of employment density on productivity.

2.2 The market border in the agglomerations

Another important issue in studying agglomeration economies is the size of the market, in other words the distance of the bank's branch that is considered significant in the spatial relationships. In spatial studies, the method of the concentric rings is used to determine different borders of the market for the particular firm. The distance is primary depending on the commuting time to the bank branch from the border.

Rosenthal and Strange (2003) concluded that, in the case of the United States, the workforce in a given industry located within a mile, is the biggest attraction for new companies, an effect that dramatically declines with distance. For the United Kingdom, Rice, Venables and Patacchini (2006) find that the productivity gains for the company are lying within the 80-minute ring.

There are discussions regarding does the distance really matter in banking industry nowadays and if matters, then what is the market area for the single branch. Kenneth et al (2008) argued that there are transportation costs for the customers to travel to the branch and information costs for the bank to obtain

information about the clients. Adams et al (2007) estimate discrete choice models of consumer choice of depository institutions where the utility an individual receives from each alternative institution is a function of the branch density of the institution in that market. They found limited substitutability between banks, therefore banks are subject to the distance measures.

Nowadays cities are in a new rapid phase of their development, in which, they are expanding and start including in their growth process other adjacent areas, forming socio-spatial entities, metropolitan areas or metropolitan regions. These include the city that generates interdependence and settlements that support these processes through mutual relations

There are studies concerned with the analysis of the productivity in the metropolitan areas. The metropolitan area requires a partnership approach in the medium and long-term development. This partnership aims at offering new opportunities for housing, investments and business that are more sustainable and consistent than a single city can provide. In addition, metropolitan areas location is attractive for academic and advanced research institutions that act in support of this development.

Metropolitan areas are argued to be efficient and coherent organization of the economic relationships in the geographical location. Therefore, metropolitan areas assumed to be the right object to study the agglomeration gains and drivers. According to Clipa (2012), metropolitan market attracts highly specialized products and services providers. Location of suppliers in geographical agglomeration is beneficial both for companies, as they are able to raise revenues due to the knowledge spillover (spatial proximity with other

suppliers), and for the clients, who benefit from easy access to a wide range of products and services by competitive prices. Increasing interaction between suppliers and buyers can provide more efficient and better timing of decisions in case of technical issues, changes of clients behavior or market structure.

Metropolitan Area border creates a labor market characterized by diversification and specialization, supported by the concentration of a variety of companies in these locations. Workforce is young, highly qualified, diverse and affordable. Workers find it useful to be in such a place where they can meet a large number of employers, because it decreases the risk of not finding a job and they are also offered increased opportunities for advancement. Investing in skills is, therefore, a priority. At the same time, employers receive a diverse, specialized, highly qualified local labor pool, which they can easily access when initiating or expanding their business (Clipa 2012).

Agglomeration economies arise as a result of knowledge spillovers that are more likely occur in urbanized metropolitan. This was also suggested by the Marshal (1920) as one of the key sources of the agglomeration. On the one hand, it creates multiple opportunities for face to face contacts that facilitate knowledge sharing. On the other hand, workers can change jobs more easily, taking with them valuable knowledge and firms are able to learn more easily from business partners.

2.3 The efficiency measure

Another issue in measuring banking branch performance is the choice of the right measure for the efficiency. Commercial banking is a very difficult service industry to measure output, technical changes, or productivity growth. First,

there is disagreement over which services banks produce and over how to measure them. Banking services are often priced implicitly through below-market interest rates on deposit balances, making observed revenue flows inaccurate measure of the performance. Banking also remains a highly regulated industry in which substantial inefficiencies have been shown to exist. As a result, technical improvements that increase the productivity of the most efficient firms may not be well reflected in the industry as a whole. A further complication is that the deposit side of banking underwent substantial deregulation in the 1980s, including the lifting of effective interest rate ceilings on certain deposits and the creation of new types of accounts. The deregulation directly raised banking costs and shifted the optimal mix between the provision of services and the payment of interest to depositors (Clipa 2012). Measurement of cost changes and productivity gains must take these factors into account, including the possibility of a period of significant disequilibrium as banks attempted to adjust to deposit deregulation.

A problem with measuring the financial flows associated with balance-sheet items, particularly loans and demand deposits, is that there are implicit revenues that cannot be easily extracted from the aggregated variables. A common measure of branch performance is budgeting, which is however criticized for its emphasis on expenses rather than profitability. Measuring the performance of a branch by its profit, which includes earnings from a wide range of services such as loans and mortgages, suffers from problems of suitably allocating revenues and expenses.

The total deposits are considered as the performance measure (Clipa 2012). Drawbacks of this simple measure are that it does not distinguish the different

kinds of deposits, which bring various profit margins. Implicit revenues currently account for over 80 percent of the revenue raised on deposits. Thus, much of the controversy regarding the treatment of deposits as an input or an output arises because the explicit revenues on deposits are relatively small. However, it is certainly one of the main business drivers of banks and easily collected and used in the statistical analysis.

The previous studies were primary focused on the study of agglomeration economies in the manufacturing and industrial sectors, as well as on the investigation of the markups and market power due to the specific spatial relationships of the banks in the financial sector. There was a little attention to the efficiency gains in the banking industry due to the concentration of the banking activity in certain locations.

Chapter 3

DATA AND INDUSTRY DESCRIPTION

3.1 Description of the industry

The distinction between spatial and non-spatial data can become the subject of extensive discussions. In general, observations for which absolute location depends on the location of other entities can be considered as spatially correlated. This is the case for the developed U.S. banking system where the competition between branches is intense. The branch location has become the part of the strategy of the bank. However, all of them compete within certain area or ring, which create a certain market place for the branches.

In order to have general idea of the distribution of the branches within the U.S., the publicly available data of the branches locations was used and mapped with the state where this branch is located. As it can be seen from the Figure 2, the concentration of branches differs from state to state, and allocation of the bank branches is not uniform. This drives to the thought of the special patterns within the industry.

The definition of the market does matter for the branch. Firstly, it allows one to define the true area within which the operations of the certain branch have observable influence. Secondly, it allows one to define whether other branches do really affect the performance of one particular branch. It is apparent that the branches in California do not significantly alter the performance of those in New York. Therefore, the correct definition of the market line lead to the adequate estimations of the interaction between customers and branches.

Here, the market is defined to be a metropolitan statistical area (MSA) since areas of such kind are working as one entity with developed inner relationships. These areas may posses the spatial relationships and allow measuring the agglomeration gains.

There are 381 metropolitan statistical areas in the continental USA. According to the 2014 estimates, New York-Newark-Jersey City (NY-NJ-PA) Metropolitan Statistical Area is the biggest one with more than 20 million of population. It can be seen from the Figure 3, that this metropolitan statistical has the greatest real financial GDP among other metropolitan areas, with financial GDP accounted on average for more than 13% of the total region real GDP – the highest share in the total GDP among other MSAs. There were more than 2600 branches in the New York MSA in 2014.

3.2 Control variables

The data about the level of the deposits in each branch is collected from the Federal Deposits Insurance Corporation database. It also includes the location of the each branch, 5-digit zip code of the area it is located, type of the bank and the services provided. The density of the branches per zip code is depicted on the Figure 4. As it can be seen, the branches are consolidated around center with some expansion to the west. Also, it can be reasonably assumed that there may exist any spatial pattern.

The unit under the research is 5-digit zip-code area. Therefore, the data about the branches performance should be aggregated. The aggregation was performed in the following way:

1) the deposits were summed up for the all branches that are located in the given zip code;
2) different locations of the banks were substituted with one singe point with longitude and latitude – geographic center of the zip code;
3) the dummy variable "Headquarter" denoted 1 if there is at least one headquarter in the given zip code;
4) the dummy variable "Service Type i" (STi) denoted 1 if there is at least one bank that provides such services in the given zip code;
5) the dummy variable "Branch Type i" (BTi) denoted 1 if there is at least one bank of such type in the given zip code.

The aggregated data described in Table 1 with translation provided in Table 2. As it can be seen, not every service type is presented in New York MSA, while every zip code has a commercial bank branch Fed member. Therefore these variables will be excluded from the further regressions and analysis, since they do not have any explanatory power.

As efficiency measure for the branch and as the dependent variable the density of deposits was taken – the amount of deposits collected per square mile. For each given zip-code the value for the land area is obtained. Therefore, the dependent and explanatory variables (except for dummy variables) are defined in terms of density, e.g. the level of the variable over the given area of the region.

Regional economic data is taken from the County Business Patterns (CBP) data published by the U.S. Census Bureau. These data are the primary information that is used to summarize location patterns in the United States and give information about the paid employees, annual income and number of

establishments. For the New York Metropolitan Area zip code level economic data is describes in Table 3.

According to the Marshall, the key source of the agglomeration for any industry is the presence of the labor force. Moreover, it required that the market of skilled labor exists what will allow company to reduce costs for the search and training of the employees. Therefore, as a proxy for the labor pools the number of the employed people in zip code is taken into the regression. Only zip codes with more than 100 employees are considered, therefore allowing for the labor to be flexible within the region.

Another important source of the agglomeration is the level of development of the services industry. Since the banking industry is purely service-oriented and also depends on other service providers such as technology or maintenance, the level of the service industry in the region may influence significantly on the density of the branches and the efficiency in deposits collection. The density of services establishments is taken as a proxy for the development of the service industry. Thus, the number of establishments per square feet is considered to be an explanatory variable for the deposits density.

As it was indicated in the World Bank Report of 2009, the income of the employees has close connection with the overall level of the economic development. Therefore, annual income was takes as a proxy for the economic development of the particular zip code. The target independent variable is the income density, e.g. the level of the income per square feet.

Overall, after selecting only meaningful entries, there are 338 observations for New York MSA. It can be seen from the density data described in Table 4 that the deposits and income density varying significantly across locations.

Chapter 4

METHODOLOGY

The data on banking firstly should be checked if there is any spatial relationships in the giving area or the branches were set randomly. The Moran I criteria allows one to see whether data obtained any spatial interrelationships. The significance of the Moran I criteria may imply but not guarantee the strong spatial interrelationship within the data. Moran I is over sensitive to the locations and the levels of the observed variables. However, the Moran I criteria significance is necessary for the further investigation of the spatial relationships in the data.

The first concern with usage of the Moran statistics is the number of the observations. The Moran's I criteria is used under the 'asymptotically normal' condition. Therefore, in order to be consistent it requires big sample, however, in the current investigation sample does not exceed 500 observations. Anselin and Florax (1995) showed that the Moran statistics also performs quite well in small samples.

The spatial relationship between the variables is modeled with the help of the spatial weighted matrix. Spatial weighted matrix designed based on the inverse distance between the regions with "minmax" normalization. Therefore, the distant zip codes have little to no influence on each other what is logical due to the common sense.

Spatial competition imposes additional restrictions on the form of the model that is taken to the data. Even with the popularity of online banking, the market for financial services such as loans and asset management is still location-specific. With non-zero transportation costs, a consumer's utility from using the services of a particular bank depends not only on the attributes of the financial services, but also on its distance from other banks. Anselin (1988) shows that the econometric consequences of failing to account for observations that are spatially dependent is the same as failing to account for autocorrelation in a time-series context. In other words, the resulting parameter estimates will be consistent, but inefficient so inferences drawn from least squares regression will be incorrect.

It is crucial to determine which specification will be used in order to describe spatial relationship in the cross-sectional data. In the absence of the large branch level datasets, the choice of the model is driven by the data available.

The spatial analysis has one inherited specific – it has endogenous variable. On the other hand, the surrounding areas influence the given area. However, at the same time it influences the surrounding areas two. Moreover, due to the complexity of the interrelation of the banking industry, there may be some unobserved variables that may be spatially correlated. Therefore, the following two models are estimated and compared based on the data available:

1) Generalized Spatial 2SLS

$$Deposits = \lambda W Deposits + \beta_1 Income + \beta_2 Services + \beta_3 Employment + \sum_{2}^{5} \beta_i BR_i + u \quad (1)$$

2) Spatial Error Model (SEM)

$$Deposits = \lambda W Deposits + \beta_1 Income + \beta_2 Services + \beta_3 Employment + \sum_{2}^{5} \beta_i BR_i + \varepsilon \quad (2)$$

$$\varepsilon = \rho \mathbf{W} \varepsilon + u$$

where: *Deposits* – logarithm of the deposits density; *Income* – logarithm of the income density; *Services* - logarithm of the service establishments density; *Employment*- logarithm of the employment density; *BRi* – branch type; *HQ* – headquarter; *W* – spatial weighted matrix.

The same approach as in Ellison and Glaeser, (1997) was taken on analysis of the spatial relationships in the industry sector. The modes (l) and (2) introduce spatial relationships and describe the underlying economic concentration but consider space in a discrete manner- organized in spatial administrative units.

Such variable as deposits and the deposits spatial lag are determined simultaneously. Therefore, the general 2-stage approach, which accounted for the simultaneous effects, was used in the first model. The second model accounts for the spatial relationships in error term. Error term is regressed against the spatial lag, allowing for some unobservable elements to be also spatially correlated in the second model.

The variables are taken in logarithms in order to investigate how the change in one variable is influenced on the change in the dependent variable, e.g. elasticity of the deposits based on the other factors.

The variable of the special interest is λ. The significant of λ indicates the presence of the special lag between the observations (bank branches), and therefore it captures spillover effect. At the same time, the services establishments, income and employment density coefficients indicate the effect of the external force on the branch performance. The dependence of the region performance from the type of the branches located and headquarters presented is showed by the coefficients in front of the branch type and headquarter.

After the estimation of the regression, the analysis of the model specification test provided based on the Akaike Information Criterion (AIC) and Ramsey RESET test, which allow one to choose the best model specification. Afterwards, the full specification of the best model (3 or 4) is estimated, in order to analyze whether there is any relationship between the deposits collection and the service types provided.

1) Generalized Spatial 2SLS

$$Deposits = \lambda W Deposits + \beta_1 Income + \beta_2 Services + \beta_3 Employment + \sum_{2}^{5} \beta_i BR_i + \sum_{11}^{30} \beta_i ST_i + u \quad (3)$$

2) Spatial Error Model (SEM)

$$Deposits = \lambda W Deposits + \beta_1 Income + \beta_2 Services + \beta_3 Employment + \sum_{2}^{5} \beta_i BR_i + \sum_{11}^{30} \beta_i ST_i + \varepsilon \quad (4)$$

$$\varepsilon = \rho \mathbf{W} \varepsilon + u$$

where: *STi* – type of services provided

Finally, the elasticity of the dependent variable with respect to the explanatory variable is calculated as:

$$\varepsilon_{deposits\ w.r.t.\ i} = \beta_i \frac{\text{Mean(deposits density)}}{\text{Mean(i)}} \quad (5)$$

Chapter 5

RESULTS

Before running the regression, the data is checked for the presence of the global spatial autocorrelation. The spatial correlation is analyzed with Moran I. With zero hypothesis of no spatial autocorrelation, the results are anticipated to be negative and the hypothesis is expected to be rejected. As discussed in Chapter 4, the rejection of zero hypotheses does not mean the presence of the strong spatial pattern within the data. However, the acceptance of the no spatial correlation at this level is definitely a sign to the absence of such relationships.

The Moran I test results for the New York MSA are depicted in the Tables 5. As it can be seen all variable have passed the test for the global spatial autocorrelation. The p-value of the Moran statistics is significant at 1% level for all variables. Therefore, it cannot be rejected that there are no spatial patterns in the data for now.

After the preliminary justification for the spatial analysis, the regression is estimated. Firstly, the estimations of the models (1) and (2) in addition to the simple OLS estimations are shown in the Table 6. The Akaike Information Criteria (AIC) and the Ramsey RESET test results are shown in the Table 7.

It can be seen from the model specification and model selection criteria that only GS2SLS has met the criteria – it has the lowest AIC at the level of 0.88 and the RESET hypothesis about the right specification is accepted (zero hypotheses – the model is specified correctly, p-value is 0.2328).

From the estimates of GS2SLS, the spatial lag of the deposits density, which is supposed to capture the spillover effects, is appeared to be significant with the coefficient of 0.381. Services density and saving association are appeared to be statistically significant factors with the coefficients of 0.999 and 0.241 respectively. The estimated elasticity of the deposits density on the explanatory variables at mean is shown in the Table 8. Branch deposits are appeared to be inelastic to the surrounding deposits density with the elasticity of 0.1136. It indicates, that the increase in the level of the deposits in the surrounding regions by 1% will be reflected in around 0.11% increase in the local deposits density.

On the other hand, the deposits density appeared to have high elasticity with respect to the services industry: the change of the services density in the location by 1% lead to the 3.3% increase in the local deposits density. It indicates on the high reliance of the local branches on the technological support and technological know-how development, what is supported by the theoretical research in this field.

The increase in the number of saving communities lead to the increase of the deposits density, what is seemed kind of obvious since the saving communities are specialized in the attracting of the deposits. However, the elasticity in the mean is negative at the level of -0.045% what is economically insignificant and can be considered as absolutely inelastic.

In comparison with GS2SLS model, OLS estimations are biased and inconsistent. The OLS predicts the employees and income density to be statistically significant in the estimation of the deposits density. There is also clear upward bias in

comparison with the GS2SLS for the labor (employment), income and services density variables.

Despite the results of the OLS are seemed to be in line with the theoretical prediction, OLS fail both model selection and model specification tests, what indicates the OLS estimates are not precise to make any conclusions.

While OLS and GS2SLS models are both have similar magnitude and the direction of the coefficients, SEM specification do not follow their pattern. The size of the coefficients and magnitude fluctuates is various directions. SEM also fails both specification tests, what indicated that the SEM model is not the right choice for the available data.

As the extension of the model, the service type variables are added to the previous specification of the OLS and GS2SLS models. The results of the regressions are depicted in the Table 9 with the specification tests in the Table 10. As was concluded earlier, only relevant service types are included in the regression, e.g. those which entities are present at least in one but not in all locations.

As in the previous case, GS2SLS pass the selection and specification test, while OLS is suspected in the omitted variables and has very high ACI of 931.94.

The upward bias of the OLS estimations is becoming large for the employment, income density and saving communities branches. In addition, OLS set the employment and income densities as statistically significant at 5% level, while GS2SLS model suggest them insignificant for the explaining the deposit density

in the region. The significance of the saving communities and development of the services industry is statistically significant at 10% and 1% respectively.

The results of the elasticity estimations are provided in the Table 11. It appeared that spillover effect became more economically significant after the adding of the service type dummies, despite the reaction still remains inelastic. The increase of the deposits density in the surrounding regions by 1% leads to the 0.13% increase in the level of the deposits density in the given location. The elasticity of the services has also rise – the 1% increase in the density of the service industry leads to the almost 3.5% increase in the deposit density. At mean, the elasticity of the deposits density is appeared to become less elastic and now constitutes -0.035% in case saving communities became by 1% more presented in the New York MSA. As it was concluded earlier, this value can be considered as economically insignificant and may not be taken into the consideration.

The service type dummies appeared to be insignificant in both specifications at 10% level, what indicated the independence of the region deposits density from the specialization or the market structure of the certain region.

Chapter 6

CONSLUSIONS

It is clear that the branch location is no longer random decision, but strategic one. Since banking sector is one of the most important sectors for the economy growth, understanding and analysis of the market structure and spatial relationship between the branches is of primary interest.

The previous research in bank branch location (Richards et al. 2008) found that the low concentration of the banks allows for the significant markups above industry average. There were also conducted several researches in the industrial sector on whether the clustering of the same companies in the urbanization location bring any effect to the productivity. Those researches found positive relationship between the proximity of the companies and productivity, however there was little research on the efficiency gains due to the spatial relationships in the banking sectors. Banking sector operates differently from the other sector; therefore the analysis of the spatial relationships in other industries is not applicable and separate research is required.

The New York Metropolitan Statistical Area (New York MSA) was investigated as the region that is oriented on financial services more than any other region in the U.S. The total deposits per square feet (deposits density) were chosen as the measure of efficiency, since deposits are the cheapest and the most affordable source of funding for the banking operations. From the analysis of the industry and previous studies, the level of income, development of the service industry,

the magnitude of the labor pool and the types of services and banks were taken as key determinants.

Based on the zip level and branch level data, the aggregated zip code level dataset was constructed. Before the actual estimation, the data was checked whether there is any chance for the spatial pattern to exist by using Moran`s statistics, which is, however, over sensitive despite necessary criteria for the further spatial analysis. Moran statistics suggests the presence of the spatial patters, as well as the distribution of the branches within the region, what is depicted in the Figure 4.

After testing three models – OLS, Spatial Error Model and Generalized Spatial 2SLS – GS2SLS model appeared to be the best fit for the spatial branch data. The two specifications of the models were estimated – the first one included the key drivers (employment, income and services) and branch type, the second – add the service types that branches provide.

In both cases the spatial lag of the deposits density, which is suggested to capture the spillover effect between the industries, is statistically significant. The elasticity of the deposit density in the given location appeared to be inelastic with 0.11% and 0.13% elasticity for the first and the second specifications respectively. At the same time the development of the service (non-finance) sector significantly affect the efficiency of the branches – the elasticity is 3.3% and 3.5% in the first and the second specifications respectively.

Branches type are in general statistically insignificant, with only saving communities appeared to be statistically significant with elasticity's -0.045% and -0.04% respectively, what is not economically significant. Therefore, it can be

concluded that branch type do not affect the density of the deposits in the given region. Neither the efficiency of the branches in the region depends on the service types provided: the coefficients appeared insignificant in both OLS and GS2SLS.

The results present the analysis of the key drivers for the deposits collection efficiency. The aforementioned approach allows one to create "what-if" simulation for the commercial banks that are looking to increase the number of branches within the region based on the unique surrounding of the target location. Specifically for the New York MSA, it allowed to estimate the overall boost in the deposits collection efficiency for the branch based on the economic development of the region and the development of service industry.

Further development of the branch performance analysis is considered in two directions. The first is the construction of the alternative measure of efficiency with accounting for the profitability of different types of deposits in the region, as well as the quality of the decision-making, which can be estimated by the overall profitability of the bank's operations. The second direction is to move from zip code level data to the branch level data. Firstly, branch level approach requires more powerful computers to process very large bulks of data, and secondly it required more detailed breakdown of the data regarding the performance of the branches and their relationships. These suggestions are expected to estimate the agglomeration effects and elasticity more precisely, however, they do not expected to change the aforementioned results significantly.

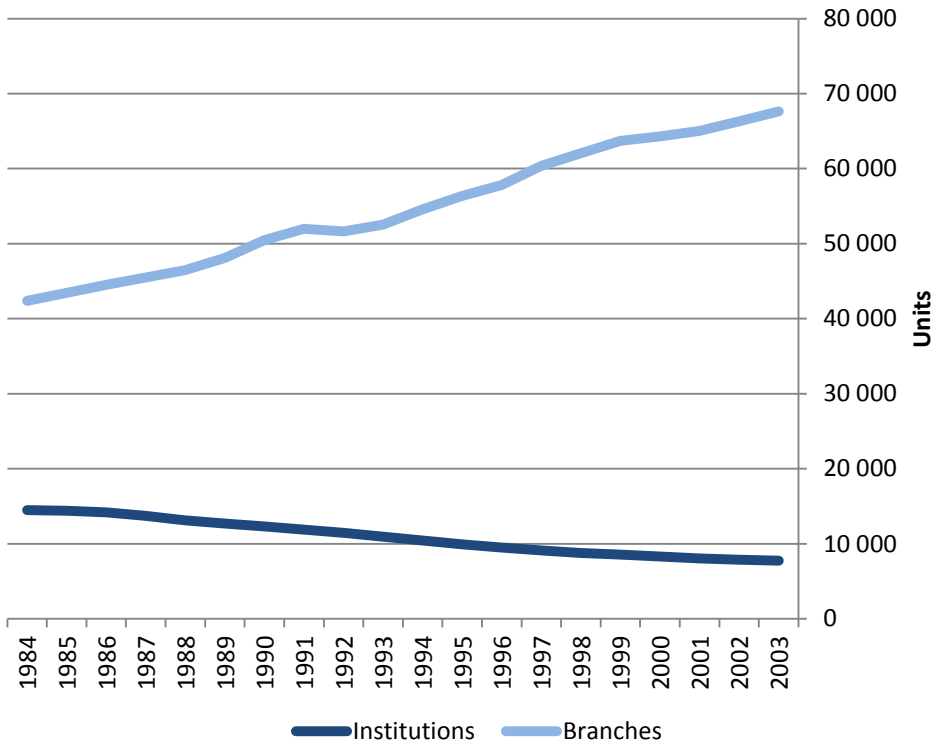

Figure 1. The development of the branching in the U.S. from 1984-2003

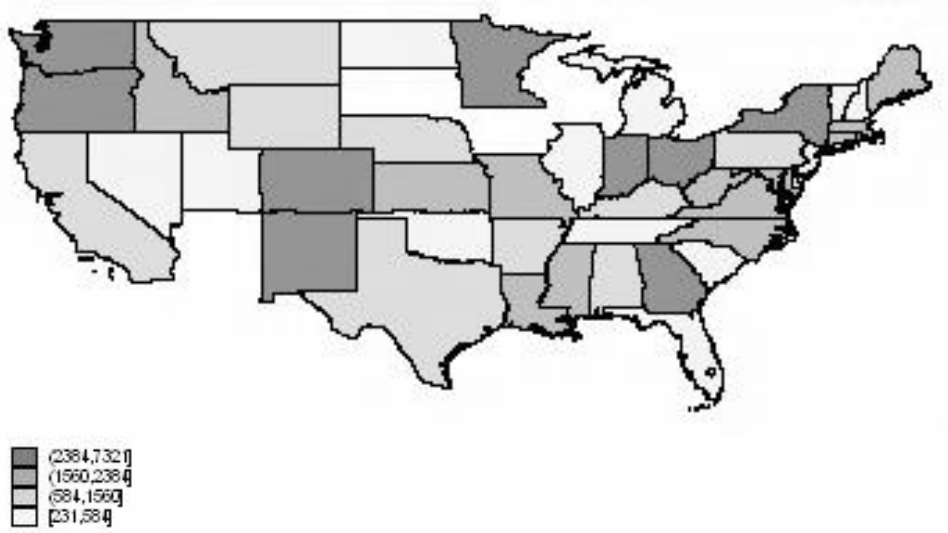

Figure 2. The distribution of the branch network in the U.S.

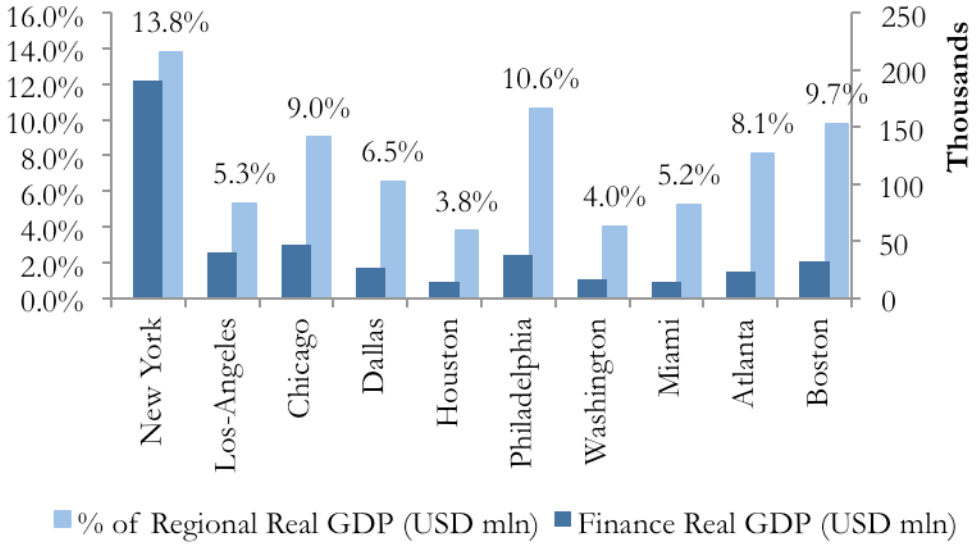

Figure 3. The level and the share of the re al financial GDP in the total real GDP for the ten largest MSA.

Figure 4. The number of branches per 5-digit zip code in the New York MSA.

Table 1. The aggregated bank branches data in New York MSA.

Variable	Mean	Std.Dev	Min	Max
Deposits	867316.1	3047379	10254	4.72E+07
HQ	0.1094675	0.3126877	0	1
BR1	1	0	1	1
BR2	0.2426036	0.4292928	0	1
BR3	0.0059172	0.0768089	0	1
BR4	0.5177515	0.5004256	0	1
BR5	0.4852071	0.5005221	0	1
BR6	0.183432	0.387594	0	1
SRV11	1	0	1	1
SRV12	0.2751479	0.4472509	0	1
SRV13	0.0118343	0.1083004	0	1
SRV14	0	0	0	0
SRV15	0	0	0	0
SRV16	0	0	0	0
SRV21	0.0384615	0.1925928	0	1
SRV22	0	0	0	0
SRV23	0.0621302	0.2417497	0	1
SRV24	0.0147929	0.1209021	0	1
SRV25	0.0059172	0.0768089	0	1
SRV26	0	0	0	0
SRV27	0.0029586	0.0543928	0	1
SRV28	0.0414201	0.1995553	0	1
SRV29	0.0325444	0.1777039	0	1
SRV30	0.0059172	0.0768089	0	1

Number of observations 338

Table 2. Definition of the services and branch types

Type	Description
BR1	Commercial bank, national (federal) charter and Fed member, supervised by the Office of the Comptroller of the Currency (OCC)
BR2	Commercial bank, state charter and Fed nonmember, supervised by the FDIC
BR3	Insured U.S. branch of a foreign chartered institution (IBA)
BR4	Savings associations, state or federal charter, supervised by the Office of Thrift Supervision (OTS)
BR5	Savings banks, state charter, supervised by the FDIC
BR6	Commercial bank, state charter and Fed member, supervised by the Federal Reserve (FRB)
ST11	Full Service Brick and Mortar Office
ST12	Full Service Retail Office
ST13	Full Service Cyber Office
ST14	Full Service Mobile Office
ST15	Full Service Home/Phone Banking
ST16	Full Service Seasonal Office
ST21	Limited Service Administrative Office
ST22	Limited Service Military Facility
ST23	Limited Service Facility Office
ST24	Limited Service Loan Production Office
ST25	Limited Service Consumer Credit Office
ST26	Limited Service Contractual Office
ST27	Limited Service Messenger Office
ST28	Limited Service Retail Office
ST29	Limited Service Mobile Office
ST30	Limited Service Trust Office
HQ	Headquarter

Table 3. The 5-digit zip code level economic data for the New York MSA.

Variable	Mean	Std. Dev.	Min	Max
Services	853.24	787.85	36.00	6,691
Employees in region	10,636.70	14,646.85	236.00	134,222
Annual income (US dollars)	630,749.30	1,514,232.00	9,956.00	19,200,000

Number of observations 338

Table 4. The description of the economic activity density in the region

Variable	Mean	Std. Dev.	Min	Max
Deposits density	11.74991	207.2861	0.0001148	3811.016
Services density	0.0005903	0.0022732	2.59E-06	0.0284883
Income density	1.679167	9.296274	0.0005336	90.83904
Labor (employee) density	0.0131324	0.0580393	0.0000137	0.5238422

Number of observations 338

Table 5. Moran I statistics for New York MSA

Variable	I statistics	p-value
Deposits density (log)	0.034	0.000
Income density (log)	0.050	0.000
Services density (log)	0.053	0.000
Labor density (log)	0.064	0.000

Number of observations 338

Table 6. Regressions estimation for the New York MSA.

Variable	GS2SLS	SEM	OLS
Income density	0.246 (0.264)	-0.112 (0.958)	0.381** (0.183)
Services establishments density	0.999*** (0.138)	-1.951** (0.783)	1.004*** (0.141)
Employees in region density	-0.297 (0.264)	1.755 (1.543)	-0.456* (0.255)
BR2 (=1 if such branch type present in region)	0.098 (0.131)	-2.218*** (0.712)	0.076 (0.133)
BR3 (=1 if such branch type present in region)	-0.497 (0.669)	-0.453 (1.481)	-0.428 (0.680)
BR4 (=1 if such branch type present in region)	0.241** (0.106)	1.048 (0.761)	0.286*** (0.105)
BR5 (=1 if such branch type present in region)	0.134 (0.105)	0.941 (0.696)	0.111 (0.107)
BR6 (=1 if such branch type present in region)	-0.141 (0.135)	2.280*** (0.758)	-0.163 (0.137)
HQ (=1 if headquarter present in region)	-0.065 (0.174)	-0.032 (0.705)	-0.052 (0.176)
Intercept	5.305*** (0.868)	-5.131 (3.777)	4.359*** (0.727)
λ	0.381* (0.199)	-0.639 (0.733)	- -
Number of observations	338	338	338

*- 10%, **- 5%, ***-1% significance level.

Table 7. Model specification test results.

Model	Akaike Information Criteria	Ramsey RESET (p-value)	Joint Significance (p-value)
GS2SLS	0.88	0.2328	0.0000
SEM	30.52	0.0000	0.0000
OLS	916.13	0.0000	0.0000

Table 8. Elasticity of the deposits density with respect to the explanatory variables at mean.

Variable	Marginal Effect	Elasticity
Spatial lag of deposits	0.3813	0.1136
Labor density	-0.2973	-0.7392
Income density	0.2459	0.2782
Services density	0.9992	3.3248
HQ	-0.0655	0.0026
BR2	0.0976	-0.0085
BR3	-0.4975	0.0011
BR4	0.2408	-0.045
BR5	0.1345	-0.0235
BR6	-0.1409	0.0093

Mean of the dependent variable = -2.7726

Table 9. Regressions estimation with service type dummies.

Variable	OLS	GS2SLS
Employment	-0.535**	-0.351
	(0.265)	(0.271)
Income	0.418**	0.262
	(0.174)	(0.184)
Services	1.057***	1.051 ***
	(0.146)	(0.141)
HQ	-0.054	-0.070
	(0.181)	(0.175)
BR2	0.039	0.062
	(0.142)	(0.137)
BR3	-0.459	-0.430
	(0.796)	(0.770)
BR4	0.2571**	0.209*
	(0.113)	(0.111)
BR5	0.093	0.123
	(0.111)	0.109
BR6	-0.178	-0.160
	(0.142)	(0.138)
ST12	0.107	0.096
	(0.125)	(0.121)
ST13	0.266	0.340
	(0.485)	(0.471)
ST21	-0.092	-0.096
	(0.282)	(0.273)
ST23	-0.008	-0.030
	(0.219)	(0.212)
ST24	0.323	0.282
	(0.453)	(0.438)
ST25	-0.265	-0.115
	(0.674)	(0.656)
ST27	-1.107	-
	(0.945)	(omitted)
ST28	0.242	0.230
	(0.273)	(0.264)
ST29	-0.114	-0.054
	(0.306)	(0.298)
ST30	-0.107	-0.272
	(0.790)	(0.768)
Intersection	4.421***	5.494***
	(0.765)	(0.903)
λ	-	0.420**
	-	(0.203)
Number of observations 338		

*- 10%, **- 5%, ***-1% significance level.

Table 10. Model specification test results with services dummies.

Model	Akaike Information Criteria	Ramsey RESET (p-value)
GS2SLS	39.24	0.0000
OLS	931.94	0.0000

Table 11. Elasticity with service type dummies.

Variable	Marginal Effect	Elasticity
Spatial lag of deposits	0.42	0.1251
Labor density	-0.3513	-0.8734
Income density	0.2623	0.2967
Services density	1.0506	3.4959
HQ	-0.0703	0.0028
BR2	0.0621	-0.0054
BR3	-0.4299	0.0009
BR4	0.2094	-0.0391
BR5	0.1227	-0.0215
BR6	-0.1598	0.0106
SRV12	0.0958	-0.0095
SRV13	0.3396	-0.0014
SRV21	-0.0965	0.0013
SRV23	-0.0302	0.0007
SRV24	0.2822	-0.0015
SRV25	-0.1145	0.0002
SRV27	-1.1901	0.0013
SRV28	0.2303	-0.0034
SRV29	-0.054	0.0006
SRV30	-0.2718	0.0006

Mean of the dependent variable = -2.772

WORKS CITED

Kenneth, Brevoort and Wolken. 2008. Does Distance Matter in Banking*? Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board*, Washington, D.C.

Clipa, Raluca Irina. 2012. Attempts to estimate the sources of agglomeration economies in IASI metropolitan area. *Anale. Seria Stiinte Economice. Timisoara*, 2012, vol. XVIII/Supplement, pages 352-358.

Brechmann, Eike and Ng Peggy. 2011. Quantifying geographical and macroeconomic effects on bank branch deposits using linear mixed models. *International Journal of Statistics and Management System*, vol. 6, no.1-2.

Berger, Allen and David Humphrey. 1992. Measurement and Efficiency Issues in Commercial Banking. *Output Measurement in the Service Sectors.* (p. 245 - 300), University of Chicago Press.

Coughlin, Cleutus, Thomas Garrett and Ruben Hernández-Murillo. 2003. Spatial Probit and the Geographic Patterns of State Lotteries. *St. Louis Federal Reserve Bank Working Paper* 2003-042B.

Richards,Timothy, Acharya Ram and Albert Kagan. 2007. Spatial Competition and Market Power in Banking. *International European Forum on Innovation and System Dynamics in Food Networks*. Austria February 15-17, 2007.

Chang, Chaudhuri and Jayaratne. 1997. Rational Herding and The Spatial Clustering of Bank Branches: An Empirical Analysis. *Federal Reserve Bank of New York Discussion Paper* Series No. 9697-24.

Krugman, Paul. 1991. Increasing Returns and Economic Geography. *Journal of Political Economy*, vol. 99, no. 3.

Hall, Appleyard. 2009. City of London, City of Learning?' Placing business education within the geographies of finance. *Journal of Economic Geography*, 9(5) 597-617.

Birkin, Clark. 2002. Using spatial models to solve difficult retail location problems, *Retail Decision Support*, Nottingham.

Petersen, Rajan. 2002 Does Distance Still Matter? The Information Revolution in Small Business Lending. *The Journal of Finance* Volume 57, Issue 6, pages 2533–2570.

Ellison, Glaeser. 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, vol. 105, no. 5.

Rousseau, Peter and Sylla Richard. 1994. Financial Systems, Economic Growth, and Globalization. *Globalization in Historical Perspective*, 1993, University of Chicago Press

Hoover, Edgar Malone. 1948. The Location of Economic Activity. New York: McGraw Hill

Marshall, Alfred. 1920. *Principles of Economics*. London: MacMillan

Anders, Malmberg, and Magnus Lagerholm. 2009. Path dependence in economic geography. *The Evolution of Path Dependence*, pp. 87–107.

Kindelberger, Charls. 1973. *The World in Depression*. Berkeley: University of California Press.

Gehrig, Thomas. 1998. Screening, cross-border banking, and the allocation of credit. *Research in Economics*, Elsevier, vol. 52(4), pages 387-407.

Keeble, David and L. Nachum. 2002. Why Do Business Service Firms Cluster? Small Consultancies, Clustering and Decentralisation in London and Southern England, *Transactions of the Institute of British Geographers*, NS, 27, 1, 1-24.

Desmet, Klaus and Esteban Rossi-Hansberg. 2010. On Spatial Dynamics. *Journal of Regional Science*, Wiley Blackwell, vol. 50(1), pages 43-63.

The World Bank. 2009. Word development report 2009: Reshaping Economic Geography. Quebecor World.

Rosenthal, Stuart and William Strange. 2003. Agglomeration, Labor Supply, and the Urban Rat Race. *Center for Policy Research*. Paper 106.

Ciccone, Antonio and Robert Hall. 1996. Productivity and the Density of Economic Activity. *The American Economic Review*, vol. 86, no. 1

Harris, Timothy and Yannis Ioannidies. 2000. Productivity and Metropolitan Density. *Discussion Papers Series, Department of Economics*, Tufts University 0016, Department of Economics, Tufts University.

Kenneth, Rogoff, and Reinhart Carmen. 2013. Banking Crises: An Equal Opportunity Menace. *Journal of Banking & Finance* Vol. 37, November, pp. 4557–573.

Adams, Robert, Kenneth Brevoort and Elizabeth Kiser. 2007. Who Competes With Whom? The Case Of Depository Institutions. *The Journal of Industrial Economics*, 55: 141–167.

Anselin, Luc and Raymond Florax. 1995. Small sample properties of tests for spatial dependence in regression models: some further results. *New Directions in Spatial Econometrics*, pp.21–74. Berlin: Springer-Verlag.