DECLARATION FOR THE PhD THESIS
To be inserted as the first page of the thesis

The undersigned

SURNAME	Bandelj
FIRST NAME	Andreja
PhD Registration Number	1287606

Thesis title:

Diversification of European and US Banks

PhD in	Finance
Cycle	23
Candidate's tutor	Prof. Andrea Sironi
Year of discussion	2012

DECLARES

Under his responsibility:

1) that, according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted, unless there is a temporary embargo in order to protect the rights of external bodies and industrial/commercial exploitation of the thesis;

3) that the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text (except in cases of a temporary embargo);

4) that in order keep the thesis on file at Biblioteca Bocconi, the University requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which

must be unalterable and that NORMADEC will indicate in each footnote the following information:

- thesis (*thesis* title)
  Diversification of European and US Banks.. ;
- by (*candidate's surname and first name)*
  Bandelj Andreja..... ;
- discussed at Università Commerciale Luigi Bocconi – Milano in (year of discussion) ...2012..... ;
- the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;
- **only in cases where another declaration has been undersigned requesting a temporary embargo**: the thesis is subject to a temporary embargo for (indicate duration of the embargo) ..... months;

5) that the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7) **choose hypothesis 7a or 7b indicated below**:

7a) that the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

Or

7b) that the thesis meets one of the temporary embargo hypotheses included in the declaration "**TEMPORARY EMBARGO REQUEST OF THE PhD THESIS**" undersigned elsewhere.

Date 27.1.2012

Signed (write first name and surname) Andreja Bandelj

Diversification of European and US banks

Dissertation in partial fulfillment of the requirements for the academic degree of Doctor of Philosophy in Finance (XXIII Cycle)

Andreja Bandelj (1287606)
Universita Commerciale Luigi Bocconi, Milan

Thesis Committee:

Prof. Andrea Sironi, Bocconi University (thesis supervisor)
Prof. Franceso Corielli, Bocconi University
Prof. Giampaolo Gabbi, University of Siena

CONTENTS

Preface

This PhD thesis consists of three chapters (working papers) in which geographic diversification of European and US banks is analyzed. The main topics of the three chapters are first briefly described below. This discussion is followed by an explanation of why banks are different from firms of other industries, implying that they should be analyzed separately.

In the first two chapters, the relationship between banks' geographic diversification and their cost of equity capital is examined. In particular the following research question is addressed: should banks diversify or should they stay focused in order to achieve lower cost of equity capital? In the first working paper (titled: *Should banks be geographically diversified? Empirical evidence from cross-country diversification of European banks*), cross-border diversification of European banks is analyzed, while in the second one (titled: *Should banks be geographically diversified? Empirical evidence from interstate diversification of US banks*), interstate diversification of US banks is addressed. The cost of equity capital is measured as an average of four commonly used models of implied cost of equity capital. A measure of geographic diversification used in this thesis is the Herfindahl-Hirschman index based on revenue dispersion across various geographic regions for the European sample and the Herfindahl-Hirschman index based on deposit dispersion across US states for the US sample of banks. The main finding of the analysis is that, other things equal, more diversified banks have a higher cost of equity capital than focused banks. This finding is consistent with agency theory, internal capital market and investors' negative reaction to this banks' business strategy.

There are several reasons why the European and the US banking industries are analyzed separately, hence in two different papers, even though these two papers both have a similar theoretical background and estimation approach. Firstly, the institutional environment in which banks from the two samples operate is different. Secondly, there are reasons to believe that the effect of interstate diversification can be similar or different from the one of cross-border diversification. On one hand, it holds that there are differences between various US states as there are differences between various European countries. States differ in their fiscal policy, in their demographic and industry structure. Based on this observation, it can be expected to see a similar effect of interstate and cross-border diversification. On the other hand, there are also a lot of characteristics that are common to all US states. They have the same language and currency, the culture is similar and moreover they share the same regulatory and supervisory structures. Based on these facts, it can be expected to see a different effect of interstate and cross-border diversification on banks' cost of equity capital. Thirdly, data availability on interstate diversification of US banks is much more detailed than the data on revenue diversification of European banks, while the data on deposit diversification of European banks is not available at all. Fourthly, previous studies analyzing

the impact of geographic diversification on banks' performance address bank cross-border diversification of European banks and interstate diversification of US banks separately.

The first part of the third paper (titled: *Clustering and matching of US banks during the crisis*) analyzes the clustering of middle-sized and large, listed and private US banks during the financial crisis which started in summer 2007. The second part investigates the relationship between geographic diversification and level and change in banks' risk during the crisis. The third part applies a matching technique to estimate causal treatment effects of a substantial increase in banks' interstate diversification before the crisis on the change in banks' risk during the crisis. The analysis measures banks' risk with Z-score and geographic diversification with the Herfindahl-Hirschman index based on deposit dispersion across US states. The conclusions of the analysis are the following. First, in general it holds that those banks that before the crisis had lower geographic diversification, lower profitability, higher Tier 1 ratio, lower ratio of mortgage loans over loans and lower level of Z-score, also had a smaller increase in risk during the crisis. Second, results indicate that the relationship between geographic diversification and the level and change in banks' risk during the crisis was not statistically significant. Third, those banks that diversified substantially before the crisis had a larger increase in risk during the crisis compared to those banks that did not diversify their operations much.

Next follows the discussion which attempts to explain why it is important to analyze financial institutions separately from firms from other industries. The fact is that banks are a specific type of firms. This is so because they have the ability to reduce the information asymmetry (adverse selection and moral hazard) problem and transaction costs. Moreover, they also have a crucial role in controlling risks; risks that are a consequence of mismatch between savers supply of deposits and investors demand for loans - maturity, counterparty, market risk etc. Consequently, it can be argued that what makes them different is the qualitative asset transformation they perform, and the fact that in the operations they use, their reputation and balance sheet items and not their limited funds (Scholtens and Van Wensveen, 2003). The banking industry is also severely regulated in order to protect depositors and preserve the safety and soundness of financial institutions. In addition, an important difference which became obvious during the recent financial crisis, between an industrial firm and a bank, is that when an industrial firm fails, other firms in that industry simply divide its market share, while in the case of a bank, its failure is not so straightforward since if a systemic institution fails this can have a domino effect on other financial institutions with which this institution operates.

Furthermore, there are differences between banks and firms from other industries, also from the perspective of financial investors. First, for potential investors, banks are more opaque than firms in other sectors (Morgan, 2002). Second, given the importance of banks for the overall economy and for maintaining the financial stability of the economy, it is appropriate

that banks are under the watchful supervision of governments and financial market regulators. Therefore hostile takeovers are a rarity in the banking industry, since the approval of the regulator is usually required before the takeover can take place. This results in the absence of a mechanism to discipline an inefficient management - market for corporate control - in the banking industry (Adams and Mehran, 2003). Lastly, it can be said that in most countries banks are subject to government guarantees on deposits; regulators also supervise banks' capital adequacy, their liquidity risk … In addition, the current crisis has again revealed that, for large banks in particular, it holds that governments will rescue them, if they get into trouble (a temporary nationalization of Fortis and the Royal Bank of Scotland are examples of this fact). All these factors can significantly reduce the shareholders' incentives to monitor banks' management (Adams and Mehran, 2003).

The facts stated above, which illustrate why banks are different from firms from other industries, influence the relationship between banks' cost of equity capital and geographic diversification, making it different from the relationship of firms from other industries. This implies that analyzing banks separately is meaningful and justified. This intuition is motivated a bit further also below. The first fact that may influence the relationship analyzed in this thesis is the one related to the cost of equity capital. Bank strict regulation, also knowing that banks will be bailed out in case of troubles induces investors to consider banks as less risky than they would consider them if this would not be so. Consequently, based on this observation, which is specific to banks and does not hold for other firms, it can be expected that banks have lower cost of equity capital than they would have without this guarantee. Another fact that may influence the relationship analyzed in this thesis is the one related to geographic diversification. This is so because economies of scale are very important for industrial firms (in order to spread fix costs of production, firms want to have as large production and be present in as many markets as possible), while for banks the decision to operate in different countries is not so straight forward. This is so because banks should also take into account other factors which are not so relevant for firms from other industries when deciding to spread their business geographically. For instance, consumers' trust in banks is so important because banking in general has a lot to do with trust – for instance, depositors need to trust banks that they will keep their savings save. It can be expected that in new markets, where banks' potential clients do not have a previous relationship with these, banks need quite some time to gain trust from new customers. Nonetheless, even factors such as different cultures and corruption count - the case of Hypo Group Alpe-Adria in Balkan region can be given as an example. Finally, banking is also an industry that generates high cash flows in favorable economic times - managers have a lot of money to spend on new investments so they might be more interested in making their banks grow than considering if these are actually investments with positive net present value.

References

Adams, R. and Mehran, H. (2003) Is corporate governance different for bank holding

companies?, *Economic Policy Review* **9**, 123–142.
Morgan, D.P. (2002) Rating banks: Risk and uncertainty in an opaque industry, *American Economic Review* **92**, 874–888.
Scholtens, B. and Van Wensveen, D. (2003), The theory of financial intermediation: an essay on what it does (not) explain, unpublished working paper.

Acknowledgments

In particular, I would like to thank my supervisor prof. Andrea Sironi, who guided me through the process of writing this dissertation. Special thanks go also to prof. Franceso Corielli and prof. Giampaolo Gabbi, the other two members of the thesis committee, and prof. Carlo A. Favero.

I would also like to thank Peter and my parents, Sonja and Aleksander, for their love and support during my studies.

Chapter 1:

Should banks be geographically diversified? Empirical evidence from cross-country diversification of European banks

1 Introduction

In the past decade we have witnessed a trend of accelerated globalization of financial institutions, with European banks being the most active players. This is the reason why this study explores this issue a bit more in detail. It links the topic of banks' geographic diversification and their cost of equity capital. By doing so, it tries to find some further evidence to answer the question of whether banks should diversify or stay focused to achieve lower cost of equity capital. Therefore, the central research question addressed is: has geographic diversification a statistically significant effect on banks' cost of equity capital? Furthermore, a hypothesis to be tested is: geographic diversification has a statistically significant effect on banks' cost of equity capital.

One of the reasons why the issue of diversification is a relevant one is because it is related to the optimum degree of diversification. There are theories supporting two opposite views concerning what is the optimal degree of diversification. On the one hand, the traditional banking theory (Diamond 1984; Boyd and Prescott 1986) which is based on a delegated monitoring argument, suggests that the optimum organization is a well diversified one (this is so because diversification reduces the cost of monitoring the borrowers). On the other hand, corporate finance theory (Jensen, 1986; Denis et al., 1997; Rajan et al., 2000) suggests that a firm should be focused in order to reduce agency problems and to maximize management's human capital. The cost of equity capital is an important variable, because it is the determinant of shareholders value. The issue raised here is that geographic diversification may lead to higher sales and earnings, but it is crucial for shareholders whether the return on invested capital exceeds the firm's cost of capital because in case where this does not hold, market value of the firm will decline. Investigating the globalization of European financial institutions is necessary because the largest European banks are the most globally active ones (see International Monetary Fund, 2007[1]).

Policymakers and regulators should find this paper interesting (whether or not banks benefit from diversification) since they affect banks by imposing regulations which create incentives

[1] According to the International Monetary Fund (2007), during 2005 the 50 largest European banks made 55% of their activities at home, 24% in the rest of the region, and 21% in the rest of the World, while the 20 largest Asian and Pacific banks made 86%, 5% and 9% of their activities at home, in the rest of the region, and in the rest of the World respectively. The 20 largest North American banks made 77%, 8% and 15% of their activities at home, in the rest of the region and in the rest of the World respectively. This is the reason why this study focuses on European banks.

either to diversify or to focus their portfolios. The imposition of capital requirements tied to banks' assets can be given as one example. Another fact that needs to be mentioned is that cost of equity capital affects the cost of raising new capital, which is important because of capital requirements. In addition, given that potential diversification effect on cost of equity capital has important implications for valuation and capital budgeting, this study can be also interesting for investors and managers of financial institutions.

Although the topic of diversification in financial institutions has been analyzed in several previous studies, the researchers did not dedicate much attention to the relationship between banks' diversification and their cost of equity capital. Another reason why this study contributes to the existing literature is because the existing studies on geographic diversification mainly focus on interstate diversification of US banks, while this one focuses on cross-border diversification of European ones. It can be added that the paper with the most similar topic to the one addressed in this paper is Waldron (2006). Waldron (2006) studies the relationship between international diversification and cost of capital among 483 S&P 500 firms for the year 2004. He obtains data on firms' cost of capital from ValuePro database (cost of equity capital is estimated by CAPM). He applies cross-sectional multiple regression analysis and finds that a higher level of market diversification was associated with a higher cost of capital for S&P 500 firms in the year 2004. This paper differs from Waldron's (2006) one in several ways. First, it applies a different measure for cost of equity capital (implied cost of equity capital, which is recommended to be used by recent empirical research analyzing the topic of cost of equity capital). Second, it considers a longer time period and not just one year. Third, it includes several control variables in the analysis. Fourth, this paper investigates cross border diversification of European banks, while Waldron (2006) investigates international diversification of S&P 500 firms. Besides, Waldron (2006) does not specify which firms he removes from the sample.

Furthermore, this paper is a matter of present interest also due to the fact that the European Commission requested that banks that have received substantial state aid after the crisis, restructure and divest a large part of their non-core activities, which includes also subsidiaries abroad (Commerzbank, Dexia, Hypo Real Estate, ING, Lloyds Banking Group, KBC, Royal Bank of Scotland, WestLB). The main finding of this analysis, which is that, other things equal, more geographic diversified banks have higher cost of equity capital than geographically focused ones, suggest that the European Commission's requests are reasonable and justified.

This paper begins by discussing the theoretical background, which can give us some intuition into why there should be a link between the variables of interest. Next, the empirical sample used and the methodology applied are described, followed by a discussion of the results. The last part concludes.

2 Theory and empirical literature review

2.1 The theory

2.1.1 Information asymmetries

There are reasons to believe that globally active firms are more difficult and costly to monitor than those that are purely domestic (Lee and Kwok, 1988) and that information asymmetries may increase the overall cost of capital. A higher degree of information asymmetries might arise because of differences across countries due to institutional, legal, language and cultural differences (see for example Burgman, 1996). Lee and Kwok (1988) also find that globally active firms have a higher agency cost than firms operating only domestically. Therefore, it can be argued that, when companies operate globally, monitoring the firm management becomes more difficult and costly, which results in higher cost of capital and less strict monitoring (the agency conflict between managers and shareholders becomes more severe). In other words, globally active firms might face higher cost of capital because of higher monitoring costs (investors have higher investigation cost and face wider information gap so they demand a higher return on their investment). Moreover, due to these higher costs not all shareholders are willing to actively monitor the manager, resulting in less strict monitoring, which might lead to suboptimal investment decisions of managers and waste of resources.

The agency theory states that managers decide for diversification to pursue their own interest at the expense of stockholders. This means that managers might have incentives to adopt and maintain a value reducing diversification strategy even if by doing so, they will reduce shareholder wealth. First, managers might be driven by empire building motives. Being in charge of large companies is associated with higher compensation, power, and prestige. Jensen (1986, 1993) states that empire-building preferences will cause managers to spend practically all available funds on investment projects. Second, given that managers have a large part of their wealth invested into the corporation they run, they are interested in diversification, because by reducing firm risk they will also reduce the risk of their individual investment portfolio (Amihud and Lev, 1981).

Some empirical studies which confirm the existence of the agency problem in banking are presented below. Bliss and Rosen (2001) analyze 32 large US banks over the period 1986-1995 and demonstrate a statistically significant relationship between the level of CEO compensation and the size of the bank. The authors conclude that CEO compensation after a major merger or rapid internal growth always increases and that this increase occurs regardless of an increase in productivity or value added. Berger and Hannan (1998) test the "quiet life" hypothesis in the banking industry - that the exercise of market power in concentrated banking markets enables banks to avoid minimizing costs. The authors estimate the cost efficiency of more than 5000 banks over the period 1980-1989 and find evidence that banks in more concentrated markets have lower cost efficiency.

Moreover, because of certain features, which are described below and are specific to banks, it can be expected that the agency conflict is particularly severe in banking. First of all, for potential investors banks are more opaque than firms in other sectors (Morgan, 2002). Second of all, in the banking industry, the market for corporate control, which is a mechanism for disciplining inefficient management, does not exist. This is a consequence of the fact that, because banks are very important for the overall economy and for maintaining the financial stability of the economy, they are under the watchful supervision of governments and financial market regulators. As a consequence hostile takeovers are a rarity in the banking industry, since the approval of the regulator is usually required before the takeover can take place. (Adams and Mehran, 2003). Third of all, the supervision of regulators can significantly reduce the shareholders' incentives to monitor banks' management. The fact is that in most countries banks are subject to government guarantees on deposits; regulators also supervise banks' capital adequacy, their liquidity risk etc. Furthermore, the current crisis has again revealed that, governments will not allow banks to fail, especially the large ones (a temporary nationalization of Fortis and Royal Bank of Scotland can be given as examples of this fact). (Adams and Mehran, 2003).

2.1.2 Risk reduction

Lewellen's (1971) financial theory of corporate diversification is based on the coinsurance effect. Lewellen (1971) argues that by combining businesses whose cash flows are less than perfectly correlated, a reduction in firms default risk can be obtained, which also serves to increase the diversified firm's debt capacity. Hann et al. (2009) develop a model with which they demonstrate that by combining firms whose cash flows are imperfectly correlated, a reduction in systematic risk can be achieved.

There is also some evidence suggesting that a diversified firm which operates in several less than perfectly correlated economies should have lower earnings volatility (Agmon and Lessard, 1977). The investors' perception of the reduced volatility as financial strength of the firm could lead to reduction in cost of equity capital. Fatemi (1984) constructs two portfolios (one consisting of internationally diversified firms while the other consisting of purely domestic firms) and compares them with respect to their performance and risk. The author does not find a statistically significant difference in the performance of the two portfolios of companies with the exception of the instance where the internationally diversified firms operate in competitive foreign markets; in this case the portfolio of the domestic firm outperforms the portfolio of internationally diversified firms. The author also concludes that the international diversification reduces the degree of systematic risk since the portfolio of internationally diversified firms has lower and more stable beta than the portfolio of undiversified firms. Nevertheless, the author also points out that at that time, the correlation between US and foreign economic activity was low.

In relation to this, it is necessary to add that the high correlation of business cycle across developed countries in the past years has become a stylized fact (see for example, Gayer, 2007, for business cycle synchronization in the euro area). Moreover, given that most of the European banks expand their activities into the rest of Europe it is clear that the co-insurance benefits might be small. Related to this argument are also the findings of Collins (1990), who concludes that there are no statistically significant differences in the risk between US firms with operations also in other developed countries and firms operating only domestically.

2.1.3 Internal capital markets

Internal capital markets might work imperfectly so that firms pursuing a geographically diversifying strategy might incur additional costs; nonetheless these firms might also internalize the capital market transactions which help them reduce their cost of capital.

The internalization theory of synergy, which was initially proposed by Caves (1971), argues that multinational firms expand abroad to internalize markets for some of their intangible assets, such as superior production skills, patents and brands, marketing abilities, managerial skills, knowledge and research or consumer goodwill. Firms invest abroad because the markets for these firm-specific assets, which firms wants to exploit, are imperfect and so the assets cannot be sold for their internal value. Through international diversification, these firms can set up a mechanism that brings buyers and sellers of these assets together within the firm. Another source of gains for internationally diversified firms comes from scale economies of these assets (the value of these assets to the firm increases with the size of the firm's activities). Or as Morck and Yeung (1991), who also provide empirical support for this theory, argue, these intangible assets have some characteristics of public goods in that their value is enhanced in direct proportion to the scale of the firm's markets.

Kogut (1983) argues that the main advantage of a multinational firm with respect to a domestic firm is the flexibility to transfer resources throughout its geographically diversified network. This means that firms can exploit market conditions by arbitraging institutional restrictions, for example, by choosing the most cost efficient location to declare profits and to raise capital. It also means that the firm is more flexible in responding to changes in international relative prices. The firm can shift the distribution to the location where demand is the highest and production to the location where production costs are the lowest. A similar line of reasoning can be found in Stein's (1997) 'winner-picking' model of investment decisions. Stein (1997) argues that under certain circumstances (headquarters are involved in funding small and focused set of projects) internal capital markets may be efficient and can create value even if they cannot relax the credit constraints firms are facing.

However, it must also be claimed that diversification can lead to inefficient cross-subsidization of less profitable business units. Researchers addressing this agency conflict between division managers and the CEO are Rajan et al. (2000), Scharfstein and Stein (2000)

and Wulf (1999). In their models, they describe division managers as rent-seeking agents who try to influence the CEO to give them higher compensation, power or resources.

The existence of internal capital markets in multinational bank holding companies was confirmed by De Hass and Van Lelyveld (2008). Authors analyze data on the 45 largest banking groups in the world and their affiliates abroad. Authors demonstrate that the parent companies trade off between lending in foreign or domestic countries and that they support weak branches abroad.

2.1.4 Investors' preferences

It is worth pointing out that it is also relevant how investors perceive geographic diversification. If investors find the strategy of geographic diversification as value reducing, or if they do not value diversification, they will require a higher return on their investment to be willing to invest in the banks' equity, which means higher cost of equity capital.

Related to this argument is the study by Morck and Yeung (1991). Authors find that investors do not value multinational firms as a mean of diversifying their portfolio internationally. Similarly, Rowland and Tesar (2004) find little evidence that multinationals increase investors' opportunities over those which are already offered by companies operating domestically. They also find that higher benefits from international diversification can be obtained by adding international assets, meaning investors can achieve higher benefits from international diversification by including into their portfolio foreign stocks rather than stocks of domestic companies, which operate also abroad. Grinblatt and Keloharju (2001) find that investors' behavior is also influenced by the firm's language, culture and distance from the investor.

Some information on the investors' response (market's expectation) to banks deciding to diversify their activities can be obtained from studies that assess the market's reaction to the announcement of merger and acquisitions of listed European banks. Cybo-Ottone and Murgia (2000) analyze large European bank merger and acquisition deals over the period 1989-1997. The authors find positive abnormal returns for the shareholders of target and acquiring banks in the case of mergers and acquisitions within the country but not in a case of mergers and acquisitions of banks from different countries. Similarly, Lepetit et al. (2004) who analyze mergers and acquisitions among European banks over the period 1991-2001 find that the most positive market reaction is in the case of transactions which are either product diversified or geographically focused. Also Beit et al. (2004) analyze large European mergers and acquisitions between 1985 and 2000. Their results indicate that the market prefers business and geographically focused transactions, meaning shareholders of the acquiring banks realize higher returns in the case of focused transactions while the target shareholders realize higher return in the case of diversified transactions. In other words, the acquiring bank has to or is willing to pay more for targets that diversify their operations. Camp and Hernando (2006)

analyze domestic and international mergers in the European financial sector over the period 1998-2002; Schmautzer (2006) analyzes international mergers and acquisitions in Europe. Both studies conclude that the abnormal returns are positive only for the target shareholders.

Based on the discussion above, it can be concluded that from a theoretical point of view, the effects of geographic diversification on cost of equity capital are controversial. Hence, this paper attempts to contribute to this debate by examining the relationship between the banks' geographic diversification and their cost of equity capital; it tries to determine if positive or negative net effects prevail.

2.2 Empirical research on the topic of banks' cost of equity capital

Zimmer and McCauley (1991) estimate the cost of capital (cost of equity capital, cost of subordinated debt) for 34 international banks from six countries over the period 1984-1990. As a proxy for cost of equity capital they use bank-level return on equity (ratio of banks' adjusted reported earnings to market capitalization). Their results show that cost of equity capital was high for banks in the United States, Canada and the United Kingdom, moderate for banks in Germany and Swiss, and low for banks in Japan.

Maccario et al. (2002) investigate whether there are country differences in the cost of equity capital of large banks from 12 developed countries over the period 1993-2001. They measure cost of equity capital by using the inverse of a bank's forward-looking price/earnings multiple. The authors conclude that both the countries' average costs of equity capital and the differences between countries' costs of equity capital have been decreasing over the sample period.

King (2009) estimates the cost of equity capital for banks in six countries over the period 1990-2009. The cost of equity capital is estimated by using the CAPM. The author concludes that cost of equity capital declined across all countries over the period 1990-2005 but then rose from 2006 onwards. The fall in the cost of equity capital can be partially explained by the decrease in risk-free rates over the period whilst the main contributor (in all countries except Japan) was the fall in the banking sector risk premium.

2.3 Empirical research on the topic of banks' geographic diversification

There are many papers written on the topic of banks' diversification; therefore, only the most relevant ones are mentioned in this section. It can be added that most of the studies are conducted on US data and provide mixed results. However, there are also some exceptions which analyze European banks. One such study is Acharya et al. (2006), which examines the effect of industrial and sector loan focus (diversification) on the return and loan risk of Italian banks. The study finds that diversification of banks' loan portfolio does not guarantee a

superior performance and/or greater safety of these banks. The authors interpret these results by suggesting that when banks expand into industries where the competition is more severe, where learning cost exists, or if managers are not skilled enough, poor monitoring incentives or greater credit risk may arise. Hayden et al. (2007) test the robustness of the results of Acharya et al. (2006) by examining a panel of portfolio diversification across different industries, economic sectors and geographic regions of German banks. They find that for the majority of their banks, diversification is negatively related to bank returns. In other words, they conclude that geographic focus has a positive effect on return for all risk profiles.

Other major studies on the topic of banks' geographic diversification are now presented in chronological order. Chong (1991) uses an event study methodology to analyze the effect of interstate banking on the risk and profitability of commercial banks. He finds that banks benefit from interstate banking because it increases their profitability; however, this increase is also associated with an increase in the banks' exposure to market risk. He argues that this might happen because diversified banks may take on more risk; they may increase their leverage and due to competitive pressures they may undertake unnecessary risk, such as risky lucrative loans.

Hughes et al. (1999) analyze the consolidation of US banks. They find that benefits come in terms of improved production efficiency and reduced risk of insolvency. However, they also find that the consolidation strategies that do not enhance interstate expansion lower the risk of insolvency, but they do not affect the efficiency.

Berger et al. (2000) estimate banking efficiency in France, Germany, Spain, the UK and the US during the 1990s. Their main conclusion that, on average, domestic banks have a higher cost and profit efficiency than foreign banks is interpreted as the home field advantage. In their subsequent paper, they analyze the effects of geographic expansion on bank efficiency for US banks. Berger and DeYoung (2001) discover that if the parent organization is efficient it can overcome any negative effects of distance by transmitting its superior knowledge to their affiliates. Therefore, their conclusion suggests that a particular optimal geographic scope for the banking organization does not exist but that the degree of geographic diversification depends on the banks' efficiency.

DeLong (2001) argues that the focus hypothesis applies to bank mergers. He analyses a sample of domestic US bank mergers and finds that (geographically and activity) focused ones create shareholder value, while those that diversify do not.

Some studies find that geographic diversification is positively associated with lending but that it is not necessarily associated with an increase in bank performance. One such study is performed by Morgan and Samolyk (2003). Authors analyze interstate diversification of US bank holding companies and find that banks' lending does increase with diversification but

that the profits of banks do not whilst the risk in their portfolio does not decrease. In contrast to these results, Akhigbe and Whyte (2003), who analyze the change in market risk of banks following the passage of Riegl-Neal Act of 1994, find that total and unsystematic risk have fallen after the passage of the act.

Some studies suggest that international diversification offers benefits because it reduces risk. Buch et al. (2004) compute the optimally diversified portfolio and compare it with banks' actual international portfolios. Their findings suggest that banks over-invest domestically, that they tend to under-invest in countries which are culturally less similar or where capital controls exist, and that there exist gains from cross-border diversification (due to the diversification of country-specific risks). Garcia-Herrero and Vázquez (2007) try to assess the potential geographic diversification gains of banks incorporated in G7 countries and Spain, which are already pursuing an internationalization strategy. They find that international banks with a larger share of assets allocated to foreign subsidiaries are able to attain higher risk-adjusted returns. Deng and Elyasiani's (2008) analysis of US banks indicates that interstate geographic diversification is associated with value premium and risk reduction, which they explain with coinsurance effect.

3 Sample

Time period: 2000–2010.

Countries included: this study uses data for listed banks from sixteen European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Main databases used:

- **Bankscope.** The Bureau Van Dijk issues the Bankscope database every month. To define the sample for every year, the December issue for that year is used.
- **Thomson Reuters Datastream** is a financial statistical database. It covers data on market indices, bonds, stocks, mutual funds, economic data, balance sheet data (**Worldscope**). Part of this database is also **I/B/E/S** (**Institutional Brokers' Estimate System**), which includes analysts' forecasts of several financial indicators: earnings per share, book value per share, cash flow per share, EDITDA per share, margin, ROA, ROE, share price among others.

Construction of the sample: The sample is constructed following these steps every year. First, listed banks from countries of interest are identified by using the December issue of Bankscope. The sample includes investment banks, commercial banks and bank holding companies.

Next, the following banks are removed:

- with total assets smaller than 10,000,000,000 euro,
  However, in order to keep the sample as balanced as possible, even those bank-years which do not satisfy the size criteria in every year but have all necessary data available, are kept in the sample. Hence, the sample even includes those banks which satisfy the size criteria only in one year, but have all other necessary data available in other years.
- with missing balance sheet data in Bankscope,
- for the purposes of calculating the cost of equity, the following conditions need to be placed: book value must be positive and median earnings' forecasts for at least the first and second year ahead must be available in I/B/E/S.

Every year banks are selected anew.

4 Methodology

4.1 Calculating cost of equity capital

The cost of equity capital can be defined as the rate of return that investors expect to make when they invest in a firm's equity. In this analysis it is estimated by calculating implied cost of equity capital. Before describing various models used, it is worthwhile reminding the reader that earlier research in finance has generally used ex post realized returns to measure the cost of equity capital. However, Elton (1999) argues that realized returns are poor proxies for expected returns whilst Fama and French (1997) show that by using CAPM and the Fama-French three-factor model, you get an imprecise proxy for a firm's cost of equity capital. Because of these problems related to ex post realized returns, recent empirical studies have started to suggest using an ex ante rate of return - the implied cost of equity capital, which is the discount rate that equates the present value of expected future cash flows to the current stock price. For example, Pástor et al. (2008) show that the implied cost of equity capital is a better measure for expected returns than realized returns. Based on these arguments, implied cost of equity capital is chosen to be used in this analysis.

Another reason why the use of implied cost of equity capital as a measure of cost of equity capital is the right one is because this measure has been applied to various analysis by several researchers. Chen et al. (2009, 2011) use the implied cost of equity capital to analyze the effect of corporate governance issue; Attig et al. (2008), Guedhami and Mishra (2009), Boubakri et al. (2010) use it to analyze ownership structure; Francis et al. (2005) use it to analyze disclosure and earnings quality, Dhaliwal et al. (2006) use it to analyze dividends and taxes, El Ghoul et al. (2011) use it to analyze tax enforcement while Hail and Leuz (2006) use it to analyze legal institutions and securities regulations. Hail and Leuz (2009) use implied cost of equity capital in an event study in which they examine cross-listings, whilst Hribar and Jenkins (2004) apply it to an event study of earnings restatements.

However, there are also some negative sides to the implied cost of equity capital. One of them is that models are based on the assumption that analysts' forecasts are an appropriate measure of investors' expectations about companies' future earnings (for example, Frankel and Lee, 1998; Easton and Sommers, 2007 show that this is not always true). Due to the limited availability of analysts' forecasts for individual companies, the number of units that can be included in the analysis is usually reduced. With respect to this point, it needs to be said that the sample includes only the largest European banks; therefore, data is available for most of them. Another problem is that for the estimation of implied cost of equity capital, expected future dividends are needed, but given that these are not directly observable, earnings forecasts are used as their proxy, which can be problematic (Cocrain, 2010).

At this point it needs to be stressed once again that implied cost of equity capital is an appropriate and suitable measure to be used in this study, because it is a widely recognized proxy for cost of equity capital, it is commonly used in the recent empirical studies (as discussed above) and also due to problems associated with estimating the cost of equity capital with ex post realized returns. Moreover, the robustness of results is checked by substituting the implied cost of equity capital with the inverse of price-earnings ratio.

Models of implied cost of equity capital

The implied cost of equity is estimated by implying four commonly used models: Claus and Thomas (2001), Gebhardt et al. (2001), Ohlson and Juettner-Nauroth (2005) as implemented in Gode and Mohanram (2003) and Easton (2004). These models are either residual income models or abnormal earnings growth models and they differ in their assumptions about growth, forecasting horizons, whether they include the impact of an industry and in the way they incorporate inflation into the final value. As the estimated cost of equity capital, the arithmetic average of these four estimates is used. By doing so, the effects of measurement errors that are associated with one particular model are mitigated. Another reason for not using one particular model is because there has yet been no consensus about which model is the best one (Botosan and Plumlee, 2005; Guay et al., 2005; Easton and Monahan, 2005). There is also no consensus among researchers about on which date to calculate the cost of equity capital; this study uses the month of June. This approach is also used in, for instance, Gebhardt et al. (2001). The reason for this choice is that by that date, market participants already receive the balance sheets' financial information for the previous fiscal year and most probably also update their expectations accordingly.

To get the inflation-adjusted cost of equity capital, inflation expectations (June issue of Consensus Forecasts) are subtracted from the nominal cost of equity estimates of each of these four estimates before calculating the average (this approach is also used in King, 2009).

Similar to other studies, observations for which the cost of equity estimates were undefined (Ohlson and Juettner-Nauroth, 2005, model) did not converge (Easton, 2004, Claus and Thomas, 2001, and Gebhardt et al., 2001, models) and had earnings growth forecasts over 200% were excluded. The cost of equity capital is calculated by employing Newton's method. The initial value of the cost of equity capital is set to 9%.

Notation used:
P_t … market price of a bank's stock at time t.
B_t … book value of equity per share at time t, $B_{t+1} = B_{t+i-1} + FEPS_{t+1} - D_{t+1}$.
$FEPS_{t+i}$ … median I/B/E/S consensus earnings per share forecast for the i-th year at time t.
POUT … banks dividend payout ratio is calculated as a bank's historical five-year average dividend payout ratio or as a current dividend payout ratio if the former is not available. The county-year median payout ratio is used if neither is available or if it is outside the range of zero and one.
g_{LT} … expected long-term growth rate is defined as the long-term forecast of annual inflation rate for a particular country as reported in April's issue of Consensus Forecasts[2].
k … cost of equity capital estimated by using the model identified in subscript.

Model 1: **Claus and Thomas (2001) (CT)** model is residual income valuation model. It says that "*the current stock price equals the current book value of equity plus the present value of future expected abnormal earnings*" (Claus and Thomas, 2001, p. 1635).

$$P_t = B_t + \sum_{i=1}^{5} \frac{FEPS_{t+i} - k_{CT} B_{t+i-1}}{(1+k_{CT})^i} + \frac{(FEPS_{t+5} - k_{CT} B_{t+4})(1+g_{LT})}{(k_{CT} - g_{LT})(1+k_{CT})^5} \quad (1)$$

I/B/E/S earnings' forecasts beyond three (or two) years are taken as reported where available; otherwise, they are generated based on the growth in FEPS1 to FEPS3 (or FEPS2 if later is not available). The long-term growth rate is set equal to long-term inflation expectations as reported in Consensus Forecasts.

Model 2: **Gebhardt, Lee and Swaminathan (2001) (GLS)** model is a residual income valuation model. The difference with respect to other models is that it uses ROE for estimating long-term earnings.

[2] For the countries: Austria, Belgium, Denmark, Finland, Greece, Ireland and Portugal, the Consensus Forecasts does not report long-term forecasts of the annual inflation rate; it only reports forecasts of the following years. Therefore for these countries (approximately 25% of sample bank years) long-term forecasts are not available. However, given that the aim of the analysis is to keep as much information as possible and that long-term forecasts are available for the majority of the sample bank years, these are kept for those countries for which they are available. Moreover, the results remain consistent if those countries which do not have long-term forecasts available are removed from the sample.

$$P_t = B_t + \sum_{i=1}^{12} \frac{FEPS_{t+i} - k_{GLS} B_{t+i-1}}{(1+k_{GLS})^i} + \frac{FEPS_{t+12} - k_{GLS} B_{t+11}}{(1+k_{GLS})^{12} k_{GLS}} \quad (2)$$

The model uses I/B/E/S forecasts for the first three years (in the case when $FEPS_{t+3}$ is not available, it uses implicit growth rate in $FEPS_{t+1}$ and $FEPS_{t+2}$ to forecast it), while it forecasts earnings after year three by assuming that ROE $_{t+3}$ declines linearly to the median industry ROE (equilibrium ROE) in the year t+T. This equilibrium ROE is measured as a historical five-year sector-specific median ROE. All private and public banks with total assets larger than 10 billion euro are classified into three groups: commercial banks, investment banks and bank holding companies as they are classified in Bankscope. Based on this classification, industry medians for each country and year are calculated. The abnormal earnings in the Tth year are assumed to be constant afterwards.

Model 3: **Ohlson and Juettner-Nauroth model (2005) (OJ)** as implemented in Gode and Mohanram (2003) is derived from the dividend discount model but it makes no restrictions on the dividend payout policy. It assumes that the short-term dividend-adjusted earnings growth rate decays asymptotically to long-term earnings growth rate, which is proxied by long-term inflation expectations as reported in Consensus Forecasts.

$$P_t = \frac{FEPS_{t+1}}{k_{OJ}} + \frac{FEPS_{t+2} - FEPS_{t+1} - k_{OJ} FEPS_{t+1}(1 - POUT)}{k_{OJ}(k_{OJ} - g_{LT})} \quad (3)$$

This model requires $FEPS_{t+1} > 0$ and $FEPS_{t+2} > 0$.

Model 4: **Easton (2004) (E)** model estimates the cost of equity capital from the modified PEG Ratio.

$$P_t = \frac{FEPS_{t+2} - FEPS_{t+1} + k_E FEPS_{t+1} POUT}{k_E^2} \quad (4)$$

This model requires $FEPS_{t+2} \geq FEPS_{t+1} > 0$.

4.2 Calculating focus (diversification)

4.2.1 Calculating geographic focus (diversification)

To measure diversification between major geographic areas in which the bank operates, the Herfindahl-Hirschman Index (HHI_{GF}), based on revenues is constructed for each bank. As HHI_{GF} rises, the bank becomes more concentrated and less diversified. In case where the bank is fully specialized (it generates all revenues in the home market) the index takes the value of 1. This is the reason why the main independent variable is named geographic focus and not geographic diversification, to avoid any misunderstandings.

$$HHI_{GF} = \left(\frac{\text{Re}v\_\text{hom}e}{REV}\right)^2 + \left(\frac{\text{Re}v\_Europe}{REV}\right)^2 + \left(\frac{\text{Re}v\_World}{REV}\right)^2$$

where Rev_home … revenues generated at home,
Rev_Europe … revenues generated in the rest of Europe,
Rev_World … revenues generated in the rest of the World,
REV … total revenues generated at home and abroad.

Companies sometimes report items such as "Eliminations", "Adjustment accounts", "Consolidation items", "Head Office/Consolidation" … these segments are dropped out of the analysis.

To calculate the above indices, Worldscope database is used (this database is also used, for example, by Joliet and Hubner, 2008). Worldscope reports geographic segment data as annual items in its database. The information is given for up to ten geographic segments of a company. If a company has more than ten geographic segments, the remaining segments are included in geographic segment 10 field. It is necessary to note that geographic regions are presented as reported by the company, which means that there is no standardization in reporting and that each company decides on its own how it reports its geographic segment data. In order to make the database more complete, banks' annual reports and banks' descriptions in Bankscope were also consulted.

So as to make geographic and business focus variables comparable and estimation results easier to interpret the ratio above is standardized:

$$Geog\_focus = \frac{HHI_{GF} - 1/3}{1 - 1/3} \quad (5)$$

This ratio takes the values from 0 to 1.

4.2.2 Calculating business focus (diversification)

This study follows the approach used in other papers (for instance, Stiroh, 2004[3]) and calculates banks' business focus (diversification) as the Herfindahl-Hirschman index:

[3] A similar approach is also applied in Acharya et al. (2006), Elsas et al. (2006) and Stiroh and Rumble (2006).

$$HHI_{BF} = \left(\frac{INT}{TOR}\right)^2 + \left(\frac{COM}{TOR}\right)^2 + \left(\frac{TRAD}{TOR}\right)^2 + \left(\frac{OTI}{TOR}\right)^2$$

where INT … net interest revenue,
COM … net commission and fee revenue,
TRAD … net trading revenue,
OTI … all other revenue,
TOR … total operating revenue, equal to the sum of the absolute values of INT, COM, TRAD and OTI.

This ratio describes the relative significance of the revenues streams. As Elsas et al. (2006) suggest absolute values are used in the calculation. The Herfindahl-Hirschman index increases with bank concentration (the more diversified the bank is, the smaller the index). When the bank is fully specialized and generates all revenues from one revenue stream, the index takes the value of 1, while when the bank is fully diversified across all four revenues streams (business areas) the index takes the value of 0.25. Similarly as in the case of geographic focus, this variable is named business focus in order to avoid any misunderstandings.

Similar to above, the ratio is standardized to make geographic and business focus variables comparable and results easier to interpret:

$$Bus\_focus = \frac{HHI_{BF} - 1/4}{1 - 1/4} \tag{6}$$

This ratio takes the values from 0 to 1.

4.3 Definition of variables

Banks' cost of equity capital is affected by firm-specific characteristics and macro level factors. In general, it can be said that the more the particular variable affects the investors' perceived riskiness of future returns, the more investors demand to be compensated for this risk, which means the higher cost of equity capital for a bank.

4.3.1 Dependent variable
Average cost of equity capital (as defined in sections 4.1)

4.3.2 Independent variable
Geographic focus (diversification) (Geog_focus) (as defined in section 4.2.1)

4.3.3 Control variables

Business focus (diversification) (Bus_focus) (as defined in section 4.2.2.)
The reason for including business focus into the regression is because Bodnar et al. (1998) show that by not including both types of focus (diversification) together into the model there emerges an omitted variable problem.

Banks' risk

Leverage (Book_lev)
Book Leverage = book value of total liabilities/book value of equity and total liabilities

A positive relationship between leverage and implied cost of equity capital is expected to be found. Expectations are based on the theory (Modigliani and Miller, 1958) and empirical evidence (Dhaliwal et al., 2006; Attig et al., 2008).

The analysis uses book and not market leverage with the aim to avoid the problem of correlation between independent variables. As is explained later, the model also includes market anomaly variables.

Credit risk (Llprov_loans)
Loan portfolio risk = Loan loss provisions/Loans

Several authors use this variable (Acharya et al., 2006; Iannotta et al., 2007) as a proxy for credit risk.

It is highly probable that investors do not know the true quality of bank loan portfolios. But they can obtain a limited amount of information on their quality on the basis of loan loss provisions made by the bank. Investors perceive the amount of loan loss provisions that management reserves to cover unexpected future losses on loan defaults as a signal of banks' credit risk (a bank making a small number of risky loans will have less loan loss provision compared to a bank taking higher risks). This signal will affect the stock market assessment of the bank's risk. A positive relationship between this variable and the cost of equity capital is expected to be found.

Market risk (Volatility)
Volatility = the standard deviation of monthly prices over the last 12 months divided by the average of the monthly price over the last 12 months, calculated in June of every year.

Given that several empirical papers (for example, Attig et al., 2008; Boubakri et al., 2008) capture firm's market risk with this variable, firm's market risk is proxied by this variable even in this paper. The authors find a positive and statistically significant relationship

between this variable and the cost of equity capital. Therefore, the expected sign of this variable is positive.

The reason why to prefer this variable over beta to proxy for banks' market risk is because recent empirical literature (for instance, Gebhardt et al., 2001) finds that beta exhibits little or no association with the implied cost of equity capital.

Information Asymmetry

The dispersion of analysts' earnings forecasts (Ln_kv)

This variable is proxied by the natural logarithm of coefficient of variation of analysts' one-year-ahead earnings per share forecasts as reported by I/B/E/S in June of every year.

Analyst forecast dispersion should be positively related to information asymmetry, which implies that there should be a positive relationship between this variable and the cost of equity capital. A positive relationship between these variables has been found, for example, in Boubakri et al. (2008) and Dhaliwal et al. (2006).

Market anomaly variables

Book to market (Bm)

Book to market = book value of equity capital/market value of equity capital.

Fama and French (1992) find a positive relationship between book-to-market equity and average return – meaning high book-to-market equity firms earn higher ex post returns than low book-to-market equity firms. They explain their findings by arguing that firms that have poor prospects according to market participants (signaled by low stock prices and high book-to-market equity) have higher expected stock returns (investors penalized them with higher costs of capital) than firms that have strong prospects.

Based on this argument, a positive relationship between this variable and implied cost of equity capital is expected to be found. A positive relationship between these variables is found by Attig et al. (2008) and Chen et al. (2009) among others.

Price momentum (Momentum)

Price momentum = a buy and hold return on the bank's stock over the period: beginning of June (t-1) until the end of May (t).

Jegadeesh and Titman (1993) find that by buying past winners and selling past losers investors could realize significant abnormal returns over the period 1965-1989. Their analysis also reveals that these returns are not due to the systematic risk of the trading strategies, while they are consistent with delayed price reactions to firm-specific information. Hence, if a price momentum is a risk proxy, it can be expected to find that high momentum stocks also have

higher implied risk premium and there should be a positive relationship between momentum variable and cost of equity capital.

However, including momentum as a control variable may be important also because it controls for the sluggishness in analysts' forecasts, meaning that revisions in investors' expectations about future earnings are immediately reflected in stock prices but they might not also be incorporated immediately into the analysts' forecasts. This implies that there might be a negative correlation between momentum and the cost of equity capital. Guay et al. (2005) suggest including recent stock returns as a control variable to control for this bias. Similar approach and finding (negative relationship between momentum and implied cost of equity capital) is also found in Chen et al. (2009).

Based on this discussion it can be argued that there might be a positive or a negative relationship between this variable and the implied cost of equity capital.

Year dummy
Year dummies are included in the analysis as control variables for common trends or business cycle effects.

Moreover, the following variables were considered for inclusion in the model: ROAA, cost-to-income ratio, natural logarithm of total assets, loans over total assets, deposits over total assets and country dummies, but given that they are not statistically significant and that including them in the model does not lead to different results, they are not included in the analysis.

4.4 Estimation methodology

4.4.1 Descriptive statistics and correlation between variables
The sample includes 110 banks. The largest number of banks is headquartered in Italy (26). Most banks are commercial banks and bank holding companies while the least number of banks is investment banks. As can be seen in Table 3, which describes the independent variables included in the analysis, banks are much more diversified across various business activities than across geographic regions. For all bank years the standardized Herfindahl-Hirschman index of business focus is on average, 0.283, while the standardized Herfindahl-Hirschman index of geographic focus is on average 0.579. It also holds that the variability of the geographic focus is higher than the variability of business focus. For all bank years the standard deviation of the standardized Herfindahl-Hirschman index of geographic focus is 0.357 while the standard deviation of the standardized Herfindahl-Hirschman index of business focus is 0.146. As expected, banks also have high leverage – average of this variable is for all banks included in the sample 0.942.

The correlation coefficients between different pairs of independent variables included in the model are the highest among the following pairs of variables: the ratio of book-to-market value of equity capital (hereafter bm) and kv (0.450), volatility and kv (0.415) (see Table 4). The correlation coefficients also suggest that the focus is negatively related to leverage, which means that more focused banks (both business as well as geographic) have a lower leverage. This could be explained by the intuition in Chong (1991), which argues that if banks want to establish their own offices abroad or take over banks in foreign countries they need to increase leverage in order to implement this business strategy. And also, by the co-insurance theory which says that because diversification reduces business risk, these banks can have higher financial risk.

Table 5 reports descriptive statistics for various models of implied cost of equity capital. As expected, there are differences between various models. The highest cost of equity capital for all bank years is calculated by the OJ model (the overall average is 0.115), while the lowest is calculated by the GLS model (the overall average is 0.084). The correlation coefficients between all pairs of different models of implied cost of equity capital are statistically significant (see Table 6). The highest correlation coefficient is between OJ and Easton model (0.997) while the lowest correlation coefficient is between OJ and GLS model (0.506).

The overall (across all countries and over the sample period) average real cost of equity capital is 10.1% (see Table 7). The country average cost of equity capital for banks from Spain, Portugal, Ireland, Sweden, Finland, the Netherlands, Norway and France is lower than the overall average cost of equity capital. While the country average cost of equity capital for banks from Denmark, Germany, Italy, the United Kingdom, Austria, Greece, Belgium and Switzerland is higher than the overall average cost of equity capital. Average cost of equity capital first increased over the period 2002–2004. The reason for this is the stock market downturn of 2002 (it can be viewed as a correction after a long bull market; this was also a period of accounting scandals: Enron, Arthur Andersen, World Com). The reason for not seeing an increase already in 2001 is because the cost of equity capital was calculated before the 11 September. The average increase in cost of equity capital from one year to another was the highest in the period 2007-2008, when it amounted to 3.32 percentage points.

The overall (across all countries and over the sample period) average of geographic focus is 0.579 (see Table 8). Banks coming from Belgium, Austria, Ireland, Switzerland, Germany, Netherlands, France, Portugal, and Sweden are more diversified than the overall average; while banks coming from Denmark, the United Kingdom, Spain, Greece, Italy, Norway and Finland are less diversified than the overall average.

4.4.2 Dynamic panel data: Blundell-Bond estimator

This paper begins investigating the relationship between geographic focus and banks' cost of equity capital by including into the fixed effects model one lag of the dependent variable as a

regressor and estimates the model by using generalized method of moments (GMM)[4]. The main theoretical reason for using dynamic panel data is because it is modeling a partial adjustment based approach. There are two commonly used dynamic panel data estimators, Arellano-Bond (1991) and Blundell-Bond (1998). As Roodman (2009) explains, the two are designed for situations with few time periods and many individuals: a linear functional relationship, a single left-hand-side variable that is dynamic (which means that it depends on its own past realizations), independent variables that are not strictly exogenous (which means that they are correlated with past and possibly current realizations of the error), fixed individual effects[5] and heteroskedasticity and autocorrelation within individuals, but not across them.

Arellano-Bond estimator (also called *difference GMM*) starts by differencing all regressors and uses the GMM. A potential weakness of the Arellano-Bond estimator is that the lagged levels are often poor instruments for first-differenced variables, especially if the variables are close to a random walk. Blundell and Bond (1998) improve the properties of the standard first-differenced GMM estimator by using additional moment conditions to obtain an estimator with improved precision and better finite-sample properties. The additional assumption is that first differences of instrumenting variables are uncorrelated with fixed effects, and it allows the introduction of more instruments and can improve efficiency. It builds a system of two equations, the original equation and the transformed one, and it is also known as *system GMM*. To put it a bit differently, the Arellano-Bond estimator uses an IV estimator based on the assumption that $E[y_{is}\Delta\varepsilon_{it}]=0$ for $s \le t-2$ in (9), so that the lags $y_{i,t-2}, y_{i,t-3},...$ can be used as instruments in the first-differenced. Blundell and Bond consider using the additional condition $E[\Delta y_{i,t-1}\varepsilon_{it}]=0$ so that they also incorporate the levels and use $\Delta y_{i,t-1}$ as instrument.

The following model is estimated:

[4] This analysis also follows Attig et al. (2008), Boubakri et al. (2008), Chen et al., (2009), Dhaliwal et al. (2006) and Hail and Leutz (2006) and investigates the relationship between geographic focus and banks' cost of equity capital by applying OLS estimator. It also follows Acharya (2006) and Hayden (2007) who use the two way fixed effects model to investigate the relationship. But given that the lag of cost of equity capital is statistically significant, it is more appropriate to use and report dynamic panel data rather than OLS or panel data estimators. Furthermore, the results are consistent if OLS and panel data estimators are used.

[5] One reason for using panel data is because it controls for unobservable firm-specific effects and bank characteristics that cannot be captured in the model and that remain constant over the investigation period. Another reason is the self-selection issue, where the central problem is that attributes that lead firms to self-select are unobserved. However, if the unobservables are constant over time the fixed effect models can control for them.

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \sum_{n=2}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (7)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus (diversification), as described in section 4.2.1;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

1-step and 2-step[6] GMM estimation (Blundell and Bond estimator) results are presented for the main model and 2-step GMM estimation (Blundell and Bond estimator) results for additional questions addressed. All the estimations are performed with the program xtabond2 in Stata (Roodman, 2006). Robust standard errors are used. For 1-step estimation, robust specifies that the standard errors are consistent in the presence of heteroskedasticity and autocorrelation within panels. In the 2-step estimation where the errors are already robust but usually downward bias, the Windmeijer's finite-sample correction for the two-step covariance matrix is calculated to correct for this.

When using Blundell and Bond estimator, two assumptions need to be tested. The first one is that the disturbances must be serially uncorrelated, which is equivalent to having no second-order serial correlation in the first-differenced residuals. By using the Hansen test of over-identifying restrictions, the second assumption, that the instruments are valid instruments - that they are uncorrelated with the first-differenced residuals, can be tested.

At this point it needs to be said that in order to lose the least data possible in this part of the analysis, the data for cost of equity capital for the year 1999 too is used. Furthermore, because of consolidation in the banking industry[7] in past years, it is reasonable that the panel is unbalanced.

[6] Two-step estimation means that the two-step estimator is calculated instead of the one-step one, that the covariance matrix is estimated by using the first-step residuals.

[7] For example, PricewaterhouseCoopers (2006) reports that in the period between 1996 and 2005 European banks spent 682 mrd euro (816 deals) acquiring banking businesses throughout the world (390 mrd euro (384 deals) of 682 mrd euro (816 deals) went for domestic deals) while Deutsche Bank (2008) reports that the number of credit institutions in the EU-15 decreased for 28% during the period from 1997 to 2006.

4.4.3 Heckman two-stage estimation technique

As additional robustness check, the Heckman's (1979) two-stage estimation technique is applied. This one allows to control for self-selection of firms that diversify (for the endogeneity of the firms' diversification decision). Li and Prabhala (2007) describe self-selection issues as a problem of having a regression that is well specified for a population, but it must be estimated using sub-samples of firms that self-select into a particular choice (C). They suggest estimating population parameters from self-selected subsamples, by firstly specifying a self-selection mechanism (a probit model in which firm i chooses E if the net benefit from doing so is positive). The system of equations can be described as:

$$C = E \equiv W_i = Z_i\gamma + \eta_i > 0 \quad (8)$$

$$C = NE \equiv W_i = Z_i\gamma + \eta_i \leq 0 \quad (9)$$

$$Y_i = X_i\beta + e_i \quad (10)$$

W_i … selection variable;

Z_i … publicly known information influencing a firm's choice;

γ … a vector of probit coefficients;

η_i … part of W_i not explained by public variables Z_i, assumed to be orthogonal to Z_i;

Y_i … dependent variable, observed only when a firm picks one of E or NE (but not both);

X_i … independent variables;

e_i … error term.

Assuming that η_i and ε_i are bivariate normal, this system can be estimated by a two-step procedure.

An approach similar to the one used in Campa and Kedia (2002) is used. Campa and Kedia (2002) analyze business diversification of US nonfinancial firms over the period 1978-1996, and find that firms self-select into becoming diversified and that self-selection explains the diversification discount.

A probit regression, with a dummy variable whether the bank is diversified as the dependent variable, is estimated in the first-step, while the choice of explanatory variables is based on variables found to influence the firms' decision to diversify in Campa and Kedia (2002). These variables are:

Industry instruments

1. Average propensity to diversify abroad = the fraction of all banks in the country, which have their operation diversified in other countries, captured by: a bank is defined as

diversified if it reports international sales larger than 0. In the analysis, listed banks are taken with total assets larger than 5 billion euro. This information is obtained in Datastream. (Industry_div)

It can be expected that the higher the fraction of geographically diversified banks in other states, the more attractive other states are for banks to diversify their operations there.

Time trends

2. Number of completed M&A transactions (financial sector only) for each country in every year over the average number of completed M&A transactions (financial sector only) for each country over the investigation period (N_deals). Database Zephyr is used to obtain this information.

   Campa and Kedia (2002) suggest capturing time trends by the existence of merger waves. It is intuitive that the more active the market for M&A, the higher the probability that a firm diversifies.

Firm specific instruments

3. Ta is total assets over average of total assets for each country in every year (in order to calculate country average listed and private banks from a country with total assets larger than 5 billion euro are considered). (Ta)
4. A number of stock exchanges the bank is listed on. (N_se)
5. A dummy variable whether the firm belongs to the Dow Jones Stoxx 600 - Europe stock index. (DJ 600)

   Campa and Kedia (2002) argue that firms are more likely to diversify if they are listed on the major exchanges (NYSE, Nasdaq, AMEX). This is so because those firms that are listed on these exchanges have a higher visibility, lower information asymmetry (they are followed by more financial analysts) and higher liquidity. According to Campa and Kedia (2002), whether a firm is included in S&P index, is a control variable for liquidity. With respect to these two variables, it needs to be said that they cannot be included in the analysis as Campa and Kedia (2002) suggest because a sample consists of European banks. Therefore, the first variable is substituted with a number of stock exchanges the bank is listed on. This variable is chosen because it can be expected that the more stock exchanges the bank is listed on, the more visible and liquid it is, and the less information asymmetry it has. The second variable is substituted with whether a bank belongs to the Dow Jones Stoxx 600 - Europe stock index, since it can be expected that this variable too can be used to control for liquidity.

   In the analysis the two variables, total assets and the number of M&A transactions in a given year, are standardized. Hence, it is important to point out that the results remain consistent even if the variables are not standardized and are used as Campa and Kedia (2002) suggest (in Campa and Kedia, 2002, the two variables are used not standardized).

The analysis is limited to the time period 2001 – 2010, because Bankscope does not report data for the last two variables (N_se and DJ 600) for the year 2000.

In the second stage, the cost of equity capital is regressed on the standardized Herfindahl-Hirschman index of geographic focus, other independent variables and the self-selection parameter (lambda).

The system can be descried as:

$$D_{it} = 1 \quad if \quad Z_i\gamma + \eta_i > 0 \tag{11}$$

$$D_{it} = 0 \quad if \quad Z_i\gamma + \eta_i \leq 0 \tag{12}$$

$$k_{avergae\ it} = d_0 + d_1 X_{it} + d_2 Geog\_focus_{it-1} + d_\lambda \lambda + \varepsilon_{it} \tag{13}$$

D_{it} … a diversification dummy equal to 1 if the firm operates in more than one geographic segment;
$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus (diversification), as described in section 4.2.1;
Z_i … a set of explanatory variables described above;
$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1;
X_{it} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);
ε_{it} … an error term.

Under the assumption that the error terms are bivariate normal, the system can be estimated as Heckman selection model.

Results

Main results of the two estimation techniques are similar and can be summed up as:

1. Cost of equity capital is influenced by the cost of equity capital of the previous year.

2. Other things being equal, more geographically focused firms have lower average cost of equity capital. Regression coefficient of geographic focus is negative and statistically significant. The regression coefficient, which is -0.012, can be interpreted as: if the standardized Herfindahl-Hirschman index of geographic focus increases by one standard

deviation, other things being equal, then expected cost of equity capital decreases by 43 basis points.

3. Regression coefficient of business focus variable is positive and statistically significant. This result implies that, other things being equal, more business focused firms have higher average cost of equity capital. Put differently, it can be said: other things being equal, those banks that have their business activities more diversified have lower average cost of equity capital. This conclusion is also consistent with findings in Campa and Kedia (2002).

4. Regression coefficients of all three measures of risk are positive: other things being equal, an increase in leverage, an increase in loan loss provisions over loans and an increase in volatility all imply an increase in cost of equity capital. However, regression coefficients of leverage and loan loss provisions over loans are not always statistical significant, while the regression coefficient of volatility is always statistically significant.

5. As expected, regression coefficients of the coefficient of variation and book-to-market ratio are positive, while regression coefficient of momentum is negative.

6. Self selection does not affect results.

Discussion of results

The theoretical interpretation of the results is the following.

Agency costs. Managers may opt for geographic diversification to pursue their own interest at the expense of shareholders. Managers might be driven by empire building motives, as being in charge of large banks is associated with higher compensation (Bliss and Rosen, 2001), power and reputation. Reputation is one of the main reasons why only a few managers will, even if it turns out that the decision to diversify their business abroad was not the right one, admit this wrong decision, and opt for disinvestment. At this point it can be added that as the bank becomes large and gains market power, managers' life becomes easier (Berger and Hannan, 1998). Furthermore, when a bank operates globally, it becomes increasingly difficult for investors to assess its true value, so they can punish such bank with higher cost of equity capital. As it was already mentioned, the problem of bank opaqueness is in general present in this industry (Morgan, 2002).

Internal capital market. One of most frequently cited reasons that companies use as an argument for a merger, are economies of scale, but on a sample of European banks, Altunbas et al. (2001) show that these exist only for small and medium-sized banks. Similar studies often interpret these results as a consequence of diseconomies of scale that occur in large

banks because of the complexity of managing such large institutions and/or the difficulties that arise when banks diversify (to a large extent) their activities geographically (Amel et al., 2004). Hence, it can be argued that large banks become too bureaucratic and therefore have difficulty in processing large amounts of information. Moreover, when managing banks operations become too complex, managers have more difficulties in controlling the organization (and the division managers), which may lead to less efficient internal control processes and duplication of costs. It is also intuitive that the more different various parts of the bank are, the more likely it is that the internal conflicts between divisional managers increase. This could lead to less efficient allocation of resources (agency conflict between division managers becomes more severe).

It is also necessary to note that when the bank expands its business to other countries, managers must successfully integrate different cultures into a homogeneous corporate culture. In their empirical study, Fiordelisi and Martelli (2010) demonstrate that corporate culture has an important role in the success of mergers and acquisitions in the European banking. In addition, it can be added that employees are a key for a successful integration. Relations with customers are heavily dependent on soft information, which is difficult to quantify. Consequently, the resignation of key executives or the emergence of morale problems due to employee turnover may lead to the loss of information. This may happen especially when the new management has little time to develop customer information (Amel et al. 2004).

Investors do not appreciate this business strategy. Investors can diversify their investment portfolio geographically themselves and they do not need a bank to do this for them. So if investors find the strategy of geographic diversification as value reducing, or if they do not value diversification, they will demand a higher return on their investment to be willing to invest in the banks' equity, which means higher cost of equity capital. As already mentioned, researchers found mixed results on the response of the cross-country mergers and acquisitions of European banks. But most often, researchers find that mergers and acquisitions are more beneficial for targets shareholders (there exists significant positive abnormal returns for shareholders of target banks) than for shareholders of acquiring banks (the researchers find either insignificant or significant and negative abnormal returns for the shareholders of acquiring banks). They also notice that the market prefers business and geographically focused transactions, which can be restated by saying that the acquiring bank has to or is willing to pay more for targets that diversify their operations. (Beit et al., 2004; Cybo-Ottone and Murgia, 2000; Lepetit et al., 2004; Schmautzer, 2006).

Furthermore, empirical studies, which have demonstrated a negative net effect of geographic diversification, explain this finding by pointing out that banks in foreign markets develop more risky credit portfolios or business practices (for instance Chong, 1991; Morgan and Samolyk, 2003). Acharya et al. (2006) explain their results with the poor monitoring

incentives or greater credit risk that may arise when banks expand into industries where the competition is more severe, where learning cost exists, or if managers are not skilled enough.

4.5 Additional questions addressed

Results of this section analysis are reported in Table 11.

4.5.1 Substituting the implied cost of equity capital with inverse price-earnings ratio as a proxy for cost of equity capital

To further check the results, an alternative model of cost of equity capital, an inverse of price-earnings ratio, is also considered. If it is assumed that next-year's earnings forecast is sufficient for valuation, cost of equity capital can be estimated as the inverse of the forward price-earnings ratio.

$$k_{EP} = \frac{FEPS_{t+1}}{P_t} \tag{14}$$

P_t … market price of a bank's common stock in June of year t;
$FEPS_{t+1}$ … I/B/E/S consensus earnings per share forecast for the next year at time t (forecasts are collected as of June of every year).

As can be seen in the Table 11 the main results do not differ much from the ones in the main section of the results.

4.5.2 Various business streams and the average cost of equity capital

In order to check if regression coefficients of different revenue streams have a different sign, the following model is estimated:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Int_{it-1} + \alpha_3 Com_{it-1} + \alpha_4 Trad_{it-1} + \\ + \sum_{n=5}^{N} \alpha_n X_{nit} + c_i + u_{it} \tag{15}$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus (diversification), as described in section 4.2.1;

Int_{it-1} … net interest revenue in total operating revenue,

Com_{it-1} … net commission and fee revenue in total operating revenue;

$Trad_{it-1}$ … net trading revenue in total operating revenue;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3.3 (banks' risk, market anomaly variables, information asymmetry variable);
c_i … individual-specific effects;
u_{it} … an error term.

As can be seen in Table 11, all regression coefficients of different revenue streams have a positive sign. In addition, it is important to add that these variables are not statistically significant even if net interest revenue over total operating revenue, net commission and fee revenue over total operating revenue and net trading revenue over total operating revenue are included in the analysis individually – one at a time.

4.5.3 Is the relationship nonlinear?
The relationship between geographic diversification and implied cost of equity capital might be nonlinear, implying that investors might find a certain degree of diversification beneficial. In order to check this intuition, the following models are estimated.

A quadratic function:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Geog\_focus^2{}_{it-1} + \sum_{n=3}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (16)$$

A cubic function:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Geog\_focus^2{}_{it-1} + \alpha_3 Geog\_focus^3{}_{it-1} + \\ + \sum_{n=4}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (17)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus (diversification), as described in section 4.2.1;

$Geog\_focus^2{}_{it-1}$ … a square of the standardized Herfindahl-Hirschman index of geographic focus (diversification);

$Geog\_focus^3{}_{it-1}$ … a cube of the standardized Herfindahl-Hirschman index of geographic focus (diversification);
X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

From Table 11 it can be seen that neither a quadratic function nor a cubic function are appropriate to explain the relationship addressed in this analysis.

4.5.4 Substituting leverage with Tier 1 or Z-score

In order to further check the robustness of the results, even the model is estimated by substituting leverage with Tier 1 and Z-score. Also in this case, the results do not change much (see Table 11), but the number of observations drops.

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \sum_{n=2}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (18)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus (diversification), as described in section 4.2.1;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

5 Conclusion

In this paper the relationship between the banks' geographic diversification and their cost of equity capital is investigated. The reason why the issue of diversification is an important one is because it is related to the optimum degree of diversification, while to investigate the globalization of European financial institutions is important because the largest European banks are the most globally active ones. The cost of equity capital is an important variable, because it is the determinant of shareholders value. Given that from a theoretical point of view the effects of geographic diversification on cost of equity capital are controversial, this paper tries to contribute to this debate by examining the relationship between the banks' geographic diversification and their cost of equity capital. It tries to determine if positive or negative net effects prevail.

In the empirical analysis, various estimation techniques are used. The sample consists of the largest public banks from sixteen European countries over the period 2000-2010. The implied cost of equity capital is estimated by implying four commonly used models (Claus and Thomas, 2001; Easton, 2004; Gebhardt et al., 2001; Ohlson and Juettner-Nauroth, 2005). To

measure diversification between main geographic areas in which the bank operates, revenue based Herfindahl-Hirschman Index is constructed.

The main finding of this analysis is that there exists a positive relationship between banks' geographic diversification and their cost of equity capital. This finding is consistent with the agency theory, internal capital market and investors' negative reaction to this banks' business strategy. Furthermore, empirical studies, which find a negative net effect of geographic diversification, explain it by pointing out that banks in foreign markets develop more risky credit portfolios or business practices, that the poor monitoring incentives may arise in these new markets, or by suggesting that managers are not skilled enough for this business strategy. It is also important to note that the majority of banks from developed countries entered new markets through acquisitions. The reasons why acquisitions rarely prove as successful in the banking industry are: the banking sector is under strong political influence, usually banks need to increase their leverage in order to finance investment (which increases the banks' financial risk), and banks usually have difficulties in fully transferring their business models to foreign subsidiaries. Moreover, developed markets are saturated and besides, entering the emerging markets is becoming increasingly difficult (the competition is becoming more severe even in emerging markets as a consequence of development of local banks) (The Economist, 2010). The successful entering can be difficult; first of all, for banks which do not have a long tradition of doing business abroad. In addition, successful entering can be difficult for those banks that did not take opportunities for acquisitions in the recent past, for example when the Latin American countries were selling ownership in banks after the debt crisis or during the privatization in Eastern Europe.

To sum up, the results of this study suggest that geographic diversification on its own is not enough for an increased market performance. It appears that investors view the strategy of geographic diversification as either value reducing or they view such banks as more risky; in either case such banks appear to have a higher cost of equity capital, which is against the goals of regulators and supervisors. Therefore, it can be said that the main result of this study may be seen as an argument against encouraging European banks to diversify their activities in various geographic countries and regions. More specifically, the minimum capital requirements risk weighting in Basel II does not take into account gains that banks obtain when they opt for geographic diversification of their business activities. Given that the results show that there is a negative relationship between geographic diversification and banks' cost of equity capital, in my opinion, the regulators should not include incentives for banks to diversify also in Basel III minimal capital requirements risk weighting. Moreover, this evidence also implies that it might be optimal to have several locally orientated banks and not only a few, highly geographically diversified large banks. Besides, such structure of the banking industry would be preferred also from a systemic risk perspective. Therefore, this paper can be concluded by saying that from a policy point of view the result suggests that

regulators should think carefully before imposing regulation which would incentivize banks to increase the level of their geographic diversification.

References

Acharya, V.V., Hasan, I. and Saunders, A. (2006) Should banks be diversified? Evidence from individual bank loan portfolios, *Journal of Business* **79**, 1355–1412.

Adams, R. and Mehran, H. (2003) Is corporate governance different for bank holding companies?, *Economic Policy Review* **9**, 123–142.

Agmon, T. and Lessard, D.R. (1977) Investor recognition of corporate international diversification, *Journal of Finance* **32**, 1049–1055.

Akhigbe, A. and Whyte, A.M. (2003) Changes in market assessment of bank risk following the Riegle-Neal Act of 1994, *Journal of Banking and Finance* **27**, 87–102.

Altunbas, Y., Gardener, E.P.M., Molyneux, P. and Moore, B. (2001) Efficiency in European banking, *European economic review* **45**, 1931–1955.

Amel, D., Barnes, C., Panetta, F. and Salleo, C. (2004) Consolidation and efficiency in the financial sector: A review of the international evidence, *Journal of Banking and Finance* **28**, 2493–2519.

Amihud, Y. and Lev, B. (1981) Risk reduction as managerial motive for conglomerate mergers, *Bell Journal of Economics* **12**, 605– 617.

Arellano, M. and Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**, 277–297.

Attig, N., Guedhami, O. and Mishra, D. (2008) Multiple large shareholders, control contest, and implied cost of equity, *Journal of Corporate Finance* **14**, 721–737.

Beitel, P., Schiereck, D. and Wahrenburg, M. (2004) Explaining M&A success in European banks, *European Financial Management* **10**, 109–139.

Berger, A.N., DeYoung, R., Genay, H. and Udell, G.F. (2000) Globalization of financial institutions: Evidence from cross-border banking performance, *Brookings-Wharton Papers on Financial Services* **3**, 23–120.

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

Berger, A.N. and Hannan, T.H. (1998) The efficiency cost of market power in the banking industry: A test of the `quiet life' and related hypotheses, *Review of Economics and Statistics* **80**, 454–465.

Bliss, R. and Rosen, R. (2001) CEO compensation and bank mergers, *Journal of Financial Economics* **61**, 107–138.

Blundell, R. and Bond, S. (1998) Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* **87**, 115–143.

Bodnar, G.M., Tang, C. and Weintrop, J. (1998), Both sides of corporate diversification: The value impacts of geographic and industrial diversification, unpublished working paper, Wharton School, University of Pennsylvania.
Boyd, J. and Prescott, E. (1986) Financial intermediary coalitions, *Journal of Economic Theory* **38**, 211–232.
Botosan, C.A. and Plumlee, M.A. (2005) Assessing alternative proxies for the expected risk premium, *The Accounting Review* **80**, 21–53.
Boubakri, N., Guedhami, O., Mishra, D. and Saffar, W. (2008), Political connections and the cost of equity capital, unpublished working paper, HEC Montreal.
Buch, C.M., Driscoll, J.C. and Ostergaard, C. (2004), Cross-border diversification in bank asset portfolios, unpublished working paper, Federal Reserve Board of Washington.
Burgman, T.A. (1996) An empirical examination of multinational corporate capital structure, *Journal of International Business Studies* **27**, 553–570.
Campa J.M. and Hernando I. (2006) M&As performance in the European financial industry, *Journal of Banking and Finance* **30**, 3367–3392.
Campa, J.M. and Kedia, S. (2002) Explaining the diversification discount, *Journal of Banking and Finance* **57**, 1731–1762.
Caves, R.E. (1971) International corporations: The industrial economics of foreign investment, *Econometrica* **38**, 1–27.
Chen, K.C.W., Chen, Z.H. and Wei, K.C.J. (2009) Legal protection of investors, corporate governance, and the cost of equity capital, *Journal of Corporate Finance* **15**, 273–289.
Chen, K.C.W., Chen, ZH. and Wei, K.C.J. (2011) Agency Costs of Free Cash Flow and the Effect of Shareholder Rights on the Implied Cost of Equity Capital, *Journal of Financial and Quantitative Analysis* **46**, 171-207.
Chong, B.S. (1991) The effects of interstate banking on commercial banks' risk and profitability, *Review of Economics and Statistics* **73**, 78–84.
Claus, J. and Thomas, J. (2001) Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* **56**, 1629–1666.
Cochrane, J.H. (2010), Discount Rates, unpublished working paper, Booth School of Business, University of Chicago.
Collins, M. (1990) A market performance comparison of U.S. firms active in domestic, developed and developing countries, *Journal of International Business Studies* **21**, 271–287.
Consensus Economics, *Consensus forecasts: A digest of international economic forecasts.*
Cybo-Ottone, A. and Murgia, M. (2000) Mergers and shareholder wealth in European banking, *Journal of Banking and Finance* **24**, 831–859.
De Haas, R. and Van Lelyveld, I. (2010) Internal capital markets and lending by multinational bank subsidiaries, *Journal of Financial Intermediation* **19**, 1–25.
DeLong, G. (2001) Stockholder gains from focusing versus diversifying bank mergers, *Journal of Financial Economics* **59**, 221–252.

Deng, S.E. and Elyasiani, E. (2008) Geographic diversification, bank holding company value, and risk, *Journal of Money, Credit and Banking* **40**, 1217–1238.
Denis, D., Denis, D. and Sarin, A. (1997) Agency problems, equity ownership, and corporate diversification, *Journal of Finance* **52**, 135–160.
Deutsche Bank (2008), European banks: The silent (r)evolution, unpublished working paper, Deutsche Bank.
Dhaliwal, D., Heitzman, S. and Li, O.Z. (2006) Taxes, leverage, and the cost of equity capital, *Journal of Accounting Research* **44**, 691–723.
Diamond, D. (1984) Financial intermediation and delegated monitoring, *Review of Economic Studies* **51**, 393–414.
Easton, P. (2004) PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital, *Accounting Review* **79**, 73–95.
Easton, P.D. and Monahan, S.J. (2005) An evaluation of accounting-based measures of expected returns, *Accounting Review* **80**, 501–538.
Easton, P. and Sommers, G. (2007) Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts, *Journal of Accounting Research* **45**, 983–1015.
Economist, The (2010, 15 May) A special report on banking in emerging markets, *Economist* **395**, 3–20.
El Ghoul, S., Guedhami, O. and Pittman, J. (2011) The role of IRS monitoring in equity pricing in public firms, *Contemporary Accounting Research* **28**, 643-674.
Elsas, R., Hackethal, A. and Holzhäuser, M. (2006), The anatomy of bank diversification, unpublished working paper, University of Munich.
Elton, E.J. (1999) Expected return, realized return, and asset pricing tests, *Journal of Finance* **54**, 1199–1220.
Fama, E. and French, K. (1992) The cross sections of expected stock returns, *Journal of Finance* **47**, 427–466.
Fama, E. and French, K. (1997) Industry costs of equity, *Journal of Financial Economics* **43**, 153–193.
Fatemi, A.M. (1984) Shareholder benefits from corporate international diversification, *Journal of Finance* **39**, 1325–1344.
Fiordelisi, F. and Martelli, D. (2010), Corporate culture and M&A results in the European banking, unpublished working paper, University of Rome.
Francis, J., Lafond, R., Olsson, P.M. and Schipper K. (2005) The market pricing of accruals quality, *Journal of Accounting and Economics* **39**, 295-327.
Frankel, R. and Lee, C. (1998) Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and Economics* **25**, 283–319.
Garcia-Herrero, A. and Vazquez, F. (2007), International diversification gains and home bias in banking, unpublished working paper, International Monetary Fund.
Gayer, C. (2007), A fresh look at business cycle synchronisation in the euro area, unpublished working paper, European Commission.

Gebhardt, W., Lee, C. and Swaminathan, B. (2001) Towards an implied cost of capital, *Journal of Accounting Research* **39**, 135–176.
Gode, D. and Mohanram, P. (2003) Inferring the cost of capital using the Ohlson-Juettner Model, *Review of Accounting Studies* **8**, 399–431.
Grinblatt, M. and Keloharju, M. (2001) How distance, language, and culture influence stockholdings and trades, *Journal of Finance* **56**, 1053–1073.
Guay, W.R., Kothari, S.P. and Shu, S. (2005), Properties of implied cost of capital using analysts' forecasts, unpublished working paper, Massachusetts Institute of Technology.
Guedhami, O. and Mishra, D. (2009) Excess control, corporate governance, and implied cost of equity: International evidence, *Financial Review* **44**, 489-524.
Hail, L. and Leuz, C. (2006) International differences in the cost of equity capital: Do legal institutions and securities regulation matter?, *Journal of Accounting Research* **44**, 485–531.
Hail, L. and Leuz, C. (2009) Cost of capital effects and changes in growth expectations around U.S. cross-listings, *Journal of Financial Economics* **93**, 428-454.
Hann R., Ogneva, M. and Ozbas, O. (2009), Corporate diversification and the cost of capital, unpublished working paper, Stanford University.
Hayden, E., Porath, D. and Westernhagen, N. (2007) Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios, *Journal of Financial Services Research* **32**, 123–140.
Heckman, J. (1979) Sample selection bias as specification error, *Econometrica* **47**, 153–161.
Hribar, P. and Jenkins, N. (2004) The effect of accounting restatements on earnings revisions and the estimated cost of capital, *Review of Accounting Studies* **9**, 337-356.
Hughes, J.P., Lang, W.W., Mester, L.J. and Moon, C.G. (1999) The dollars and sense of bank consolidation, *Journal of Banking and Finance* **23**, 291–324.
Iannotta, G., Nocera, G. and Sironi, A. (2007) Ownership structure, risk, and performance in the European banking industry, *Journal of Banking and Finance* **31**, 2127–2149.
International Monetary Fund (2007) *Global financial stability report* **April**.
Jegadeesh, N. and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* **48**, 65–91.
Jensen, M.C. (1986) Agency costs of free cash flow, corporate finance and takeovers, *American Economic Review* **76**, 323–329.
Jensen, M.C. (1993) The modern industrial revolution, exit, and the failure of internal control systems, *Journal of Finance* **48**, 831–880.
Joliet, R. and Hubner, G. (2008) Corporate international diversification and the cost of equity: European evidence, *Journal of International Money and Finance* **27**, 102–123.
King, M.R. (2009), The cost of equity for global banks: A CAPM perspective from 1990 to 2009, unpublished working paper, Bank for International Settlements.
Kogut, B. (1983) *Foreign Direct Investment as a Sequential Process,* The MIT Press, Cambridge, Massachusetts.

Lee, K.C. and Kwok, C.C.Y. (1988) Multinational corporations vs. domestic corporations: International environmental factors and determinants of capital structure, *Journal of International Business Studies* **19**, 195–217.
Lepetit, L., Patry, S. and Rous, P. (2004) Diversification versus specialization: An event study of M&As in the European banking industry, *Applied Financial Economics* **14**, 663–669.
Lewellen, W. (1971) A pure financial rationale for the conglomerate merger, *Journal of Finance* **26**, 521–537.
Li, K. and Prabhala, N.R. (2007) Self-selection models in corporate finance, in: E. Eckbo (ed.), *Handbook of Corporate Finance: Empirical Corporate Finance*, Vol. **1**, Elsevier, Amsterdam, pp. 37–86.
Maccario, A., Sironi, A. and Zazzara, C. (2002), Is banks' cost of equity capital different across countries? Evidence from the G10 countries major banks, unpublished working paper, Libera Università Internazionale degli Studi Sociali Guido Carli.
Modigliani, F. and Miller, M.H. (1958) The cost of capital, corporation finance, and the theory of investment, *American Economic Review* **48**, 261–297.
Morck, R. and Yeung, B. (1991) Why investors value multinationality, *Journal of Business* **64**, 165–186.
Morgan, D.P. (2002) Rating banks: Risk and uncertainty in an opaque industry, *American Economic Review* **92**, 874–888.
Morgan, D.P. and Samolyk, K. (2003), Geographic diversification in banking and its implications for bank portfolio choice and performance, unpublished working paper, Bank for International Settlements.
Ohlson, J. and Juettner-Nauroth, B. (2005) Expected EPS and EPS growth as determinants of value, *Review of Accounting Studies* **10**, 349–365.
Pastor, L., Sinha, M. and Swaminathan, B. (2008) Estimating the intertemporal risk-return tradeoff using the implied cost of capital, *Journal of Finance* **63**, 2859–2897.
Rajan, R., Servaes, H. and Zingales, L. (2000) The cost of diversity: The diversification discount and inefficient investment, *Journal of Finance* **55**, 35−80.
Roodman, D. (2009) How to do xtabond2: An introduction to difference and system GMM in Stata, *The Stata Journal* **9**, 86–136.
Rowland, P.F. and Tesar, L.L. (2004) Multinationals and the gains from international diversification, *Review of Economic Dynamics* **7**, 789–826.
PricewaterhouseCoopers (2006), European banking consolidation, unpublished working paper, PricewaterhouseCoopers.
Scharfstein, D.S. and Stein, J.C. (2000) The dark side of internal capital markets: Divisional rent-seeking and inefficient investment, *Journal of Finance* **55**, 2537−2564.
Schmautzer, D. (2006), Cross-border bank mergers: Who gains and why?, unpublished working paper, University of Muenster.
Schmid, M.M. and Walter, I. (2009) Do financial conglomerates create or destroy economic value?, *Journal of Financial Intermediation* **18**, 193–216.

Stein, J.C. (1997) Internal capital markets and the competition for corporate resources, *Journal of Finance* **52**, 111–133.
Stiroh, K.J. (2004) Do community banks benefit from diversification?, *Journal of Financial Services Research* **25**, 135–160.
Stiroh, K.J. and Rumble, A. (2006) The darkside of diversification: The case of U.S. financial holding companies, *Journal of Banking and Finance* **30**, 2131–2161.
Waldron, D.G. (2006) International diversification and the cost of capital: Is more necessarily better?, *International Business & Economics Research Journal* **5**, 23-38.
Wulf, J. (1999), Influence and inefficiency in the internal capital market: Theory and evidence, unpublished working paper, University of Pennsylvania.
Zimmer, S.A. and McCauley, R.N. (1991) Bank cost of capital and international competition, *Federal reserve bank of New York quarterly review*, 33–59.

Tables

Table 1: Banks' cost of equity capital estimated by researchers

	Zimmer & McCauley (1991)	Maccario et al. (2002)	King (2009)	
	1984-1990	1993-2001	1993-2001	2002-2009
France		7.7%	10.6%	7.3%
Germany	6.9%	7.0%	11.4%	9.0%
United Kingdom	9.8%	8.9%	9.5%	6.6%
Japan	3.1%	2.8%	12.0%	11.2%
Canada	10.3%	12.0%	10.7%	5.4%
United States	11.9%	8.8%	10.4%	7.2%
Switzerland	5.3%	8.2%		
Belgium		8.9%		
Spain		8.0%		
Italy		7.6%		
Netherland		9.0%		
Sweden		9.7%		

Source: King (2009), Maccario et al. (2002), Zimmer & McCauley (1991)

Table 2: Number of banks included in the sample by country of incorporation

Country	Number of banks
Austria	3
Belgium	4
Denmark	4
Finland	2
France	7
Germany	10
Greece	7
Ireland	5
Italy	26
Netherlands	4
Norway	2
Portugal	5
Spain	9
Sweden	3
Switzerland	6
United Kingdom	13
Total	110

Table 3: Descriptive statistics for independent variables

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
Geog_focus	0.577	0.579	0.357	0.231	1.000	807
Bus_focus	0.267	0.283	0.146	0.181	0.361	807
Leverage	0.947	0.942	0.038	0.931	0.961	807
Llprov_loans	0.005	0.005	0.005	0.002	0.007	807
Ln_kv	-2.431	-2.351	0.835	-2.904	-1.860	807
Momentum	0.070	0.054	0.329	-0.171	0.258	807
Bm	0.555	0.718	0.519	0.412	0.823	807
Volatility	0.099	0.120	0.085	0.069	0.146	807
Total_assets	85,400,000	253,000,000	376,000,000	26,400,000	313,000,000	807

Notes: Total assets are in 1000 euro. Geog_focus is geographic focus; its calculation is described in section 4.2.1. Bus_focus is business focus; its calculation is described in section 4.2.2. Leverage is the ratio between book value of total liabilities and sum of book value of equity and total liabilities. Llprov_loans is the ratio between loan loss provisions and loans. Ln_kv is a natural logarithm of the coefficient of variation of analysts' one-year-ahead earnings per share forecasts as reported by I/B/E/S. Momentum is a buy and hold return on the bank's stock over the period: beginning of June (t-1) until the end of May (t). Bm is the ratio of book value of equity capital to market value of equity capital. Volatility is the standard deviation of monthly prices over the last 12 moths divided by the average of the monthly price over the last 12 moths.

Table 4: Correlation coefficients between independent variables

Variable	Geog_ focus	Bus_ focus	Leverage	Llprov _loans	Ln_kv	Momentum	Bm	Volatility
Bus_focus	0.323***	1						
Leverage	-0.130***	-0.269***	1					
Llprov_loans	0.038	0.109***	-0.065*	1				
Ln_kv	0.007	-0.004	-0.053	0.307***	1			
Momentum	0.013	-0.077**	-0.006**	-0.112***	-0.378***	1		
Bm	0.063*	0.063*	-0.030	0.119***	0.450***	-0.363***	1	
Volatility	-0.080**	0.065*	0.022	0.100***	0.415***	-0.337***	0.388***	1
Ln_total _assets	-0.592***	-0.476***	0.416***	0.033	0.028	-0.084**	0.033	-0.013

Notes: Variables are defined in Table 3.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 5: Descriptive statistics for implied cost of equity capital models

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
CT	0.087	0.094	0.036	0.073	0.105	807
GLS	0.079	0.084	0.035	0.060	0.099	807
OJ	0.104	0.115	0.047	0.086	0.132	807
Easton	0.100	0.112	0.046	0.082	0.128	807
Average	0.092	0.101	0.037	0.079	0.112	807

Notes: CT is the cost of equity capital, estimated by the Claus and Thomas (2001) model. GLS is the cost of equity capital estimated by Gebhardt et al. (2001) model. OJ is the cost of equity capital estimated by Ohlson and Juettner Nauroth (2005) model. Easton is cost of equity capital estimated by Easton (2004) model. Average the average cost of equity capital, calculated as the average of these four models. The calculation of variables is defined in Section 4.1.

Table 6: Correlation coefficients between different models of implied cost of equity capital

	CT	GLS	OJ	Easton
CT	1			
GLS	0.630***	1		
OJ	0.742***	0.506***	1	
Easton	0.766***	0.542***	0.997***	1
Average	0.879***	0.730***	0.943***	0.958***

Notes: Variables are defined in Table 5.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 7: Mean implied cost of equity capital and geographic focus (diversification) by year

Year	Mean of Average	Mean of Geog_focus
2000	0.084	0.617
2001	0.090	0.593
2002	0.097	0.617
2003	0.095	0.617
2004	0.099	0.584
2005	0.087	0.578
2006	0.088	0.577
2007	0.088	0.584
2008	0.121	0.551
2009	0.146	0.507
2010	0.157	0.490
Average	0.101	0.579

Table 8: Mean implied cost of equity capital and geographic focus (diversification) by banks' country of incorporation

Country	Mean of Average	Mean of Geog_focus
Austria	0.106	0.284
Belgium	0.111	0.230
Denmark	0.102	0.640
Finland	0.097	0.937
France	0.101	0.417
Germany	0.102	0.391
Greece	0.109	0.774
Ireland	0.091	0.300
Italy	0.103	0.787
Netherlands	0.099	0.407
Norway	0.100	0.839
Portugal	0.090	0.450
Spain	0.079	0.770
Sweden	0.092	0.515
Switzerland	0.124	0.349
United Kingdom	0.103	0.655
Average	0.101	0.579

Table 9: Estimated dynamic panel data model (Blundell-Bond dynamic panel data estimator): geographic focus and banks' cost of equity capital

		DPD 1 step	DPD 2 step
Variables		**(1)**	**(2)**
L.Average	**b**	**0.312*****	**0.327*****
	se	0.086	0.095
Geog_focus	**b**	**-0.015*****	**-0.012*****
	se	0.005	0.004
Bus_focus	**b**	**0.026***	**0.024****
	se	0.014	0.012
Leverage	**b**	**0.086*****	**0.083*****
	se	0.017	0.022
Llprov_loans	**b**	**0.115**	**0.182**
	se	0.305	0.319
Ln_kv	**b**	**0.008*****	**0.007*****
	se	0.002	0.002
Momentum	**b**	**-0.018*****	**-0.019*****
	se	0.004	0.004
Bm	**b**	**0.015*****	**0.014*****
	se	0.004	0.004
Volatility	**b**	**0.041****	**0.043***
	se	0.019	0.023
	constant	**0.023**	**0.018**
	se	0.019	0.024
	N	741	741
	Number of groups	110	110
	Number of instruments	84	84
	Wald chi2	940.27	713.83
	z(AR1)	-4.19	-4.38
	z(AR2)	1.66	1.65
	Hansen J test	70.13	70.13

Notes: Variables are defined in Table 3. Dependent variable is the cost of equity capital. b refers to the regression coefficient, se to robust standard error. Instruments: all lags of Average as the GMM instruments and year dummies and explanatory variables as IV instruments.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 10: Estimated Heckman selection model: geographic focus and banks' cost of equity capital

Variables		(1)
Geog_focus	**b**	**-0.025*** **
	se	0.006
Bus_focus	**b**	**0.028*** **
	se	0.010
Leverage	**b**	**0.026**
	se	0.123
Llprov_loans	**b**	**0.421**
	se	0.282
Ln_kv	**b**	**0.009*** **
	se	0.002
Momentum	**b**	**-0.012** **
	se	0.005
Bm	**b**	**0.018*** **
	se	0.004
Volatility	**b**	**0.078*** **
	se	0.023
	constant	**0.068**
	se	0.117
Ta	**b**	**0.412*** **
	se	0.058
Industry_div	**b**	**1.248*** **
	se	0.309
DJ 600	**b**	**0.110**
	se	0.151
N_se	**b**	**0.082*** **
	se	0.021
N_deals	**b**	**-0.002**
	se	0.001
	constant	**-1.093*** **
		0.205
	Lambda	0.020***
	N	729
	Censored N	172

Notes: Variables are defined in Table 3. In addition, Ta is total assets over country average of total assets, Industry_div is the fraction of banks in the country, which are geographically diversified, DJ 600 is a dummy variable whether the firm belongs to the Dow Jones Stoxx 600 - Europe stock index, N_se is a number of stock exchanges the bank it is listed on, N_deals is the number of completed M&A transactions in a given year over the average number of M&A transactions in each country over the sample period. b refers to the regression coefficient, se standard error estimated by Jackknife option. ***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 11: Additional questions addressed

Variables		(1)	(2)	(3)	(4)	(5)	(6)	(7)
L.Average	**b**		**0.338*** **	**0.323*** **	**0.319*** **	**0.264*** **	**0.317*** **	**0.322*** **
	se		0.081	0.093	0.094	0.093	0.086	0.085
L.ep	**b**	**0.415*** **						
	se	0.089						
Geog_focus	**b**	**-0.009*** **	**-0.009*** **	**0.005**	**-0.029**	**-0.014*** **	**-0.014*** **	**-0.014*** **
	se	0.003	0.004	0.014	0.037	0.004	0.004	0.005
Int	**b**		**0.015**					
	se		0.012					
Com	**b**		**0.030**					
	se		0.027					
Trad	**b**		**0.019**					
	se		0.012					
Geog_focus^2	**b**			**-0.016**	**0.069**			
	se			0.012	0.088			
Geog_focus^3	**b**				**-0.054**			
	se				0.056			
Tier 1	**b**					**0.057**		
	se					0.063		
Zscore$_{mv}$	**b**						**-0.001**	
	se						0.001	
Zscore$_{bv}$	**b**							**0.000**
	se							0.001
Bus_focus	**b**	**0.001**		**0.025** **	**0.025** **	**0.022***	**0.019**	**0.018**
	se	0.012		0.013	0.013	0.012	0.012	0.012
Leverage	**b**	**0.084** **	**0.099*** **	**0.087*** **	**0.089*** **			
	se	0.037	0.026	0.024	0.026			

Cont. of Table 11

Llprov_loans	**b**	**-0.465**	**0.314**	**0.143**	**0.165**	**0.328**	**0.230**	**0.269**
	se	0.291	0.262	0.322	0.340	0.305	0.312	0.309
Ln_kv	**b**	**-0.002**	**0.007*****	**0.007*****	**0.007*****	**0.007*****	**0.007*****	**0.007*****
	se	0.001	0.002	0.002	0.002	0.002	0.002	0.002
Momentum	**b**	**-0.029*****	**-0.019*****	**-0.019*****	**-0.019*****	**-0.019*****	**-0.020*****	**-0.020*****
	se	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Bm	**b**	**0.004**	**0.014*****	**0.014*****	**0.014*****	**0.014*****	**0.013*****	**0.015*****
	se	0.003	0.004	0.004	0.004	0.004	0.004	0.004
Volatility	**b**	**-0.012**	**0.040***	**0.044***	**0.044***	**0.048****	**0.046****	**0.042***
	se	0.017	0.024	0.024	0.023	0.024	0.023	0.024
	constant	**-0.018**	**-0.012**	**0.012**	**0.014**	**0.099*****	**0.100**	**0.095*****
	se	0.034	0.031	0.026	0.026	0.015	0.015	0.014
	N	741	741	741	741	655	714	712
	N. of groups	110	110	110	110	102	110	109
	N. of instruments	84	86	85	86	84	84	84
	Wald chi2	568.14	728.62	658.47	683.16	610.24	677.98	660.71
	z(AR1)	-4.00	-4.63	-4.38	-4.34	-3.91	-4.31	-4.34
	z(AR2)	0.37	1.90	1.94	1.9	2.01	1.70	1.71
	Hansen J test	69.42	72.46	71.6	71.27	66.47	70.71	71.9

Notes: Variables are defined in Table 3. Ep is earnings per share forecast for the next year over market price of a bank's common stock in June of year t. Int is net interest revenue in total operating revenue. Com is net commission and fee revenue in total operating revenue. Trad is net trading revenue in total operating revenue. Tier 1 is tier 1 capital over total risk-weighted assets. $Zscore_{mv}$ is equal to the sum of the average return on assets (calculated by using five years of past data) and the ratio of market value of equity capital in total assets; over the standard deviation of return on assets (calculated by using five years of past data). $Zscore_{bv}$ is equal to the sum of the average return on assets (calculated by using five years of past data) and the ratio of book value of equity capital in total assets; over the standard deviation of return on assets (calculated by using five years of past data). Dependent variable is the cost of equity capital. b refers to the regression coefficient, se to robust standard error. Estimation method: 2-step GMM estimates. Instruments: all lags of Average as the GMM instruments and year dummies and explanatory variables as IV instruments. ***/**/* denotes statistical significance at the 1%/5%/10% level.

Chapter 2:

Should banks be geographically diversified? Empirical evidence from interstate diversification of US banks

1 Introduction

Interstate diversification of American banks was initially constrained by severe regulation. However, over the years this regulation was weakened and the barriers, which disable banks to expand geographically, were reduced. Therefore, since the passage of Riegle-Neal Interstate Banking and Branching Efficiency Act in 1994, which gave the states the option to permit interstate branching, we have witnessed increasing trend in geographic diversification of American banks. Banks have been spreading their operations across many markets within the US. In relation to this fact, the question of whether interstate geographic diversification has a statistically significant effect on banks' cost of equity capital has been left unanswered. Therefore, this study addresses this issue and by linking the topic of banks' geographic diversification and their cost of equity capital, it tries to find some further evidence to answer the question of whether banks should diversify or stay focused to achieve lower cost of equity capital.

The reason why the issue of diversification is a relevant one is because it is related to the optimum degree of diversification. There are theories supporting two opposite views concerning what is the optimal degree of diversification. On one hand, the traditional banking theory (Diamond, 1984; Boyd and Prescott, 1986) suggests that the optimum organization is a well diversified one, while, on the other hand, corporate finance theory (Jensen, 1986; Denis et al., 1997; Rajan et al., 2000) suggests that a firm should be focused in order to reduce agency problems and to maximize management's human capital. The cost of equity capital is an important variable, because it is a determinant of shareholders' value. It is necessary to stress that geographic diversification may lead to higher sales and earnings; yet, for shareholders it is crucial whether the return on invested capital exceeds the firm's cost of capital because in the case when this does not hold market value of the firm will decline and the cost of capital will increase.

This analysis can be interesting for policymakers and regulators since they affect banks by imposing regulations which create incentives either to diversify or to focus their portfolios. As an example, it can be given the imposition of capital requirements tied to banks' assets. Furthermore, another important fact that needs to be mentioned is that cost of equity capital affects the cost of raising new capital, which is important because of capital requirements. The results have important implications also for financial investors and managers of financial institutions, since potential diversification effect on cost of equity capital has important implications for valuation and capital budgeting.

This paper contributes to the existing literature in several ways. It is the first paper that addresses the relationship between banks' interstate diversification and their cost of equity capital. Another characteristic of this paper in relation to other studies on bank diversification is that it jointly takes into account geographic and business diversification while most of the previous researchers focus only on one type of diversification. Furthermore, both theoretical views and empirical findings about the effects of interstate geographic diversification on various measures of banks' performance and risk, are mixed and inconclusive. Consequently, this analysis tries to contribute to this debate by examining the relationship between the banks' geographic diversification and their cost of equity capital, it tries to determine if positive or negative net effects prevail.

The paper with the most similar topic to the one addressed in this one is Waldron (2006). By applying cross-sectional multiple regression analysis, he finds that a higher level of market diversification was associated with a higher cost of capital for S&P 500 firms in the year 2004. This paper differs from Waldron's (2006) paper in several ways. It investigates interstate diversification of US banks, while Waldron (2006) investigates international diversification of S&P 500 firms. It applies a different measure for cost of equity capital (it applies implied cost of equity capital, which is recommended to be used by recent empirical research analyzing the topic of cost of equity capital, while Waldron (2006) obtains firms' cost of capital from ValuePro database, in which cost of equity capital is estimated by CAPM). It also uses a panel of data and not just one year as Waldron (2006) does. This paper includes several control variables while Waldron (2006) does not. Waldron (2006) also does not specify which firms he removes from the sample.

This paper begins by describing the theory, which can give some intuition why there should be a link between the variables of interest. It continues by describing the methodology and empirical sample used. In the main part of the analysis it presents the results whereas in the last section, it concludes.

2 Theory and empirical literature review

A theoretical basis for the expected existence of a relationship between banks' geographic diversification and their cost of equity capital can be found in theories related to the information asymmetry, risk reduction and internal capital market.

First, it can be said that, when companies operate globally, monitoring the firm management becomes more complex and costly (this is due to a higher degree of information asymmetries), which results in higher cost of capital and less strict monitoring (the agency conflict between managers and shareholders becomes more severe) (Burgman, 1996; Lee and Kwok, 1988). Furthermore, managers might decide for diversification to pursue their own interest at the expense of stockholders (Amihud and Lev, 1981; Jensen, 1986, 1993).

Empirical studies, which confirm the existence of the agency problem in banking, are, for example, Bliss and Rosen (2001) and Berger and Hannan (1998). In addition, because of certain features, which are specific to banks, it can be expected that the agency conflict is particularly severe in banking. Firstly, for potential investors banks are more opaque than firms in other sectors (Morgan, 2002). Secondly, given the importance of banks for the overall economy and for maintaining the financial stability of the economy, it is reasonable that banks are under the watchful supervision of governments and financial market regulators. This also means that hostile takeovers are a rarity in the banking industry, since the approval of the regulator is usually required before the takeover can take place. This results in the absence of a mechanism to discipline an inefficient management - market for corporate control - in the banking industry (Adams and Mehran, 2003). Lastly, it can be said that in most countries banks are subject to government guarantees on deposits; regulators also supervise banks' capital adequacy, their liquidity risk … Additionally, the current crisis has again revealed that governments will rescue large banks in particular, if they get into trouble. All these factors can significantly reduce the shareholders' incentives to monitor banks' management (Adams and Mehran, 2003).

Second, Lewellen's (1971) financial theory of corporate diversification is based on the coinsurance effect. Lewellen argues that by combining businesses whose cash flows are less than perfectly correlated, the reduction in firms' default risk can be obtained. Similar reasoning can be found also in Agmon and Lessard (1977), Fatemi (1984), Hann et al. (2009).

Third, internal capital markets might work imperfectly so firms pursuing geographically diversifying strategies might incur additional costs. Hence, on one hand, diversification can lead to inefficient cross-subsidization of less profitable business units (Rajan et al., 2000; Scharfstein and Stein, 2000; Wulf, 1999). On the other hand, these firms might also internalize the capital market transactions, which help them reduce their cost of capital (Caves, 1971; Kogut, 1983; Stein, 1997). The existence of internal capital markets in multinational bank holding companies was confirmed by De Hass and Van Lelyveld (2008), who demonstrate that the parent companies trade off between lending in foreign or domestic country and that they support weak branches abroad.

Lastly, investors' perception of this banks' strategy is also important. If investors find the strategy of geographic diversification as value reducing, or if they do not value it, they will demand a higher return on their investment to be willing to invest in the banks' equity, which implies higher cost of equity capital. Morck and Yeung (1991) find that investors do not value multinational firms as a means of diversifying their portfolio internationally. Rowland and Tesar (2004) find little evidence that multinationals increase investors' opportunities over those which are already offered by companies operating domestically.

Given the previous discussion, it can be argued that, from a theoretical point of view, the effects of geographic diversification on the cost of equity capital are controversial. Hence, this study examines this relationship with the aim to see if positive or negative net effects prevail.

Testable hypothesis: interstate diversification has a statistically significant effect on banks' cost of equity capital.

The topic of banks' cost of equity capital did not receive much attention from researchers. So far, only some empirical papers have been written on this issue (King, 2009; Maccario et al., 2002; Zimmer and McCauley, 1991). While King (2009) and Zimmer and McCauley (1991) focus mainly on estimating the cost of equity capital, Maccario et al. (2002) also investigate whether there are country differences in the cost of equity capital of large banks from 12 developed countries over the period 1993-2001. Researchers use various estimation techniques to estimate cost of equity capital.

The topic of banks' geographic diversification received much more attention than the topic of banks' cost of equity capital. Most of the studies are conducted on the sample of US banks and provide mixed results (Akhigbe and Whyte, 2003; Berger et al., 2000; Berger and DeYoung, 2001; Chong, 1991; DeLong, 2001; Garcia-Herrero & Vázquez, 2007; Hughes et al., 1999; Morgan and Samolyk, 2003). Acharya et al. (2006) and Hayden et al. (2007) are the most relevant studies on this topic which analyze European banks.

3 Sample

Sample includes US banks over the period 1995-2010.

The main databases used in the analysis are the following:

- **Bankscope.** The Bureau Van Dijk issues the Bankscope database every month. To define the sample for every year, the December issue for that year is used.
- **Thomson Reuters Datastream** is a financial statistical database. It covers data on market indices, bonds, stocks, mutual funds, economic data, balance sheet data (**Worldscope**). Part of this database is also **I/B/E/S** (**Institutional Brokers' Estimate System)**, which includes analysts' forecasts of several financial indicators: earnings per share, book value per share, cash flow per share, EDITDA per share, margin, ROA, ROE, share price among others.
- **The Summary of Deposits** (SOD) is the Federal Deposit Insurance Corporation (FDIC) database. It is an annual survey of branch office deposits for all FDIC-insured institutions as of June 30. This survey has been conducted since 1934, but on-line data is available from 1994 afterwards.

Construction of the sample: The sample is constructed following these steps every year. First, listed US banks are identified by using the December issue of Bankscope. Next, the following banks are removed:

- with total assets smaller than 10,000,000,000 US dollars,
  Nevertheless, in order to keep the sample as balanced as possible, also those bank-years, which do not satisfy the size criteria in every year, but have all necessary data available, are kept in the sample,
- which are not included into the SOD database,
- with missing balance sheet data in Bankscope,
- for the purposes of calculating the cost of equity, the following conditions need to be placed: book value must be positive and median earnings forecasts for at least the first and second year ahead must be available in I/B/E/S.

Every year banks are selected anew.

4 Methodology

4.1 Cost of equity capital

The cost of equity capital can be defined as the rate of return that investors expect to make when they invest in a firm's equity. In this analysis it is estimated by implied cost of equity capital. While earlier research in finance has used ex post realized returns to measure the cost of equity capital, recent research has demonstrated problems related to ex post realized returns (Elton, 1999; Fama and French, 1997). This is the reason why recent empirical studies have started to suggest using an ex ante rate of return (for instance: Pástor et al., 2008) - the implied cost of equity capital, which is the discount rate that equates the present value of expected future cash flows to the current stock price.

Below are listed various empirical analysis to which researchers apply implied cost of equity capital as a measure of cost of equity capital too. Chen et al. (2009, 2011) use the implied cost of equity capital to analyze the effect of corporate governance issue; Attig et al. (2008), Guedhami and Mishra (2009) and Boubakri et al. (2010) use it to analyze ownership structure, Francis et al. (2005) use it to analyze disclosure and earnings quality, Dhaliwal et al. (2006) use it to analyze dividends and taxes, El Ghoul et al. (2011) use it to analyze tax enforcement while Hail and Leuz (2006) use it to analyze legal institutions and securities regulations. Hail and Leuz (2009) use this measure in an event study in which they examine cross-listings whilst Hirbar and Jenkins (2004) apply it to an event study of earnings restatements.

To sum up, the decision to rely on implied cost of equity capital as a measure for cost of equity capital in this analysis is based on finding that this is a better measure for cost of equity

capital than measures which are based on realized returns. Moreover, this measure is also widely used in recent corporate finance literature.

Nevertheless there are also some problems related to using implied cost of equity capital as a measure for cost of equity capital. First, these models are based on the assumption that analysts' forecasts are an appropriate measure of investors' expectations about companies' future earnings, Frankel and Lee (1998), Easton and Sommers (2007) show that this is not always true. Second, there might be limited availability of analysts' forecasts for individual companies. However, with respect to this point it needs to be said that in this analysis this is not problematic because the sample consists of large public banks for which data is usually available. Third, expected future dividends are needed but since these are not directly observable, earnings forecasts are used as their proxy, which can be problematic (Cocrain, 2010).

At this point it needs to be stressed once again that implied cost of equity capital is an appropriate and suitable measure to be used in this study[8] because it is a widely recognized proxy for cost of equity capital and it is commonly used in the recent empirical studies. In addition, estimating the cost of equity capital with ex post realized returns has its problems. Furthermore, in the last section of the analysis, the robustness of results is checked by substituting the implied cost of equity capital with the inverse of price-earnings ratio.

The implied cost of equity is estimated by implying four commonly used models: Claus and Thomas (2001), Gebhardt et al. (2001), Ohlson and Juettner-Nauroth (2005) as implemented in Gode and Mohanram (2003) and Easton (2004). These models are either residual income models or abnormal earnings growth models. As the estimated cost of equity capital, the arithmetic average of these four estimates is used. Reasons for doing so are the following. First, consensus among researchers about which model is the best one to be used has not been reached yet (Botosan and Plumlee, 2005; Guay et al., 2005; Easton and Monahan, 2005). Second, by not using one particular model, the effects of measurement errors that are associated with a particular model can be avoided.

Similarly as in Gebhardt et al. (2001) cost of equity capital is calculated in June every year. The reason for doing so is that by this date, market participants already receive the balance sheets financial information for the previous fiscal year and most probably also update their expectations accordingly. To get the inflation-adjusted cost of equity capital, inflation expectations (June issue of Consensus Forecasts) are subtracted from the nominal cost of equity estimates of each of these four estimates before calculating the average (this approach

[8] Even though FED suggests using CAPM to estimate the cost of equity capital, this study uses implied cost of equity capital models, because it aims at having comparable results between the European and the US sample, and also because of the problems already discussed associated with estimating cost of equity capital with CAPM.

is also used in King, 2009). Similarly to other studies, this one too excludes observations for which the cost of equity estimates were undefined (Ohlson and Juettner-Nauroth, 2005, model), did not converge (Easton, 2004, Claus and Thomas, 2001, and Gebhardt et al., 2001, models) and had earnings growth forecasts over 200%. The cost of equity capital is calculated by employing Newton's method. The initial value of the cost of equity capital is set to 9%.

Models of implied cost of equity capital

Notation used:
P_t … market price of a bank's stock at time t.
B_t … book value of equity per share at time t, $B_{t+1} = B_{t+i-1} + FEPS_{t+1} - D_{t+1}$.
$FEPS_{t+i}$ … median I/B/E/S consensus earnings per share forecast for the i-th year at time t.
POUT … banks dividend payout ratio is calculated as a bank's historical five-year average dividend payout ratio or as a current dividend payout ratio if the former is not available. The county-year median payout ratio is used if neither is available or if it is outside the range of zero and one.
g_{LT} … expected long-term growth rate is defined as the long-term forecast of annual inflation rate as reported in April issue of Consensus Forecasts.
k … cost of equity capital estimated by using the model identified in subscript.

Model 1: **Claus and Thomas (2001) (CT)** model is residual income valuation model:

$$P_t = B_t + \sum_{i=1}^{5} \frac{FEPS_{t+i} - k_{CT} B_{t+i-1}}{(1+k_{CT})^i} + \frac{(FEPS_{t+5} - k_{CT} B_{t+4})(1+g_{LT})}{(k_{CT} - g_{LT})(1+k_{CT})^5} \quad (1)$$

I/B/E/S earnings forecasts beyond three (or two) years are taken as reported where available. Otherwise, they are generated based on the growth in $FEPS_1$ to $FEPS_3$ (or $FEPS_2$ if latter is not available). The long-term growth rate is set equal to long-term inflation expectations as reported in Consensus Forecasts.

Model 2: **Gebhardt, Lee and Swaminathan (2001) (GLS)** model is a residual income valuation model. The difference with respect to other models is that it uses ROE for estimating long-term earnings.

$$P_t = B_t + \sum_{i=1}^{12} \frac{FEPS_{t+i} - k_{GLS} B_{t+i-1}}{(1+k_{GLS})^i} + \frac{FEPS_{t+12} - k_{GLS} B_{t+11}}{(1+k_{GLS})^{12} k_{GLS}} \quad (2)$$

The model uses I/B/E/S forecasts for the first three years (in the case where $FEPS_{t+3}$ is not available, it uses implicit growth rate in $FEPS_{t+1}$ and $FEPS_{t+2}$ to forecast it), while it forecasts earnings after year three by assuming that ROE_{t+3} declines linearly to the median industry

ROE (equilibrium ROE) in the year t+T. This equilibrium ROE is measured as a historical five-year sector-specific median ROE. All private and public commercial banks and bank holding companies with total assets larger than 10 billion US dollars are classified in two groups: commercial banks and bank holding companies as they are classified in Bankscope. Based on this classification industry medians for every year is calculated. The abnormal earnings in the Tth year are assumed to be constant afterwards.

Model 3: **Ohlson and Juettner-Nauroth model (2005) (OJ)** as implemented in Gode and Mohanram (2003) is derived from the dividend discount model, but it makes no restrictions on the dividend payout policy. It assumes that the short-term dividend-adjusted earnings growth rate decays asymptotically to long-term earnings growth rate, which is proxied by long-term inflation expectations as reported in Consensus Forecasts.

$$P_t = \frac{FEPS_{t+1}}{k_{OJ}} + \frac{FEPS_{t+2} - FEPS_{t+1} - k_{OJ} FEPS_{t+1}(1 - POUT)}{k_{OJ}(k_{OJ} - g_{LT})} \tag{3}$$

This model requires $FEPS_{t+1} > 0$ and $FEPS_{t+2} > 0$.

Model 4: **Easton (2004) (E)** model estimates the cost of equity capital from the modified PEG Ratio.

$$P_t = \frac{FEPS_{t+2} - FEPS_{t+1} + k_E FEPS_{t+1} POUT}{k_E^2} \tag{4}$$

This model requires $FEPS_{t+2} \geq FEPS_{t+1} > 0$.

4.2 Definition of focus (diversification)

4.2.1 Geographic focus (diversification)

For each bank, a deposit based Herfindahl-Hirschman Index (HHI_{GF}), which measures banks' geographic focus, is constructed. This index is the sum of squared deposit shares in each state where the bank operates and it decreases with banks' diversification – the more diversified the bank is, the smaller the index. The index takes the value of 1 for those banks that are fully specialized. This is also the reason why the main independent variable is called geographic focus and not geographic diversification.

$$HHI_{GF} = \sum_{i=1}^{50} \left(\frac{Deposits_{State\ i}}{Total\ \ Deposits} \right)^2$$

The information on deposit distribution across various states of the US is obtained from the Federal Deposit Insurance Corporation (FDIC) the summary of deposits (SOD) database. This database was already used before by various researchers (Morgan and Samolik, 2003; Deng and Elyasiani, 2008). SOD reports geographic segment data as annual items in its database. Data contains information about branch office deposits for all FDIC-insured institutions as of June 30. Data is available on-line from 1994 afterwards. The information is given for all federal countries in which the bank has its branches. The reporting is standardized.

In order to make geographic and business focus variables comparable and estimation results easier to interpret, the ratio above is standardized:

$$Geog\_focus = \frac{HHI_{GF} - 1/50}{1 - 1/50} \quad (5)$$

This ratio takes the values between 0 and 1.

At this point it is worthwhile to point out that the main difference with respect to the variables used in this and the previous paper of the thesis is the way geographic diversification is measured. In the previous paper, the measure of geographic diversification is based on revenues generated in various worlds' regions (at home, in the rest of Europe and in the rest of the World), while in this paper much more detailed measure of geographic diversification can be calculated, as the Summary of Deposits database reports exact banks' deposit dispersion across US states. This means that the variable used in the European sample is a flow variable while the variable used in the US sample is a stock variable. And it might be expected that a stock variable is more stable over time than a flow variable and that a higher level of diversification might result if revenues are used. Some remarks can be made vis-à-vis this problem. First, this is a problem that arises because of data availability and it cannot be overcome. Moreover, this study follows other studies in using these databases to collect information about cross-country diversification of European banks (Datastream) and interstate-diversification of US banks (Summary of Deposits database). Second, another evidence, which does not reject the use of two different measures of geographic diversification, is the observation related to comparing within variation (which is a variation over time) and between variation (which is a variation across banks) of geographic focus variable for European and US banks. For the whole sample period between variation of geographic focus variable is 0.34 for European banks, which is a bit more than 0.29 for US banks. While within variation of geographic focus variable is 0.14 for US banks and 0.11 for European banks, and more importantly if only the time period 2000-2010 is considered, the difference between the two variations decreases even further (between variation of geographic focus variable for US banks increases to 0.30 and within variation drops to 0.10). Hence, it

can be argued that because variations of two measures of geographic diversification do not differ much between the two samples, the use of two different measures of geographic diversification can be appropriate. Third, it is important to add that there is also a positive side of having a measure of geographic diversification based on two different variables (flow and stock) because this implies that the results are robust to using both types of measure.

4.2.2 Business focus (diversification)

This study follows other researchers (Acharya et al., 2006; Elsas et al., 2006; Stiroh, 2004; Stiroh and Rumble, 2006) and calculates banks' business focus (diversification) as the Herfindahl-Hirschman index:

$$HHI_{BF} = \left(\frac{INT}{TOR}\right)^2 + \left(\frac{COM}{TOR}\right)^2 + \left(\frac{TRAD}{TOR}\right)^2 + \left(\frac{OTI}{TOR}\right)^2$$

where INT … net interest revenue,
COM … net commission and fee revenue,
TRAD … net trading revenue,
OTI … all other revenue,
TOR … total operating revenue, equal to the sum of the absolute values of INT, COM, TRAD and OTI.

This ratio describes the relative significance of the revenues streams. As Elsas et al. (2006) suggest absolute values are used in the calculation. The Herfindahl-Hirschman index decreases with bank diversification – the more diversified the bank is, the smaller the index. The index takes the value of 1 for those banks that are fully specialized, hence for those banks that generate all revenues from one revenue stream. Similarly as is done for the geographic focus variable even this one is called business focus in order to avoid any misunderstandings.

Similarly as above, the ratio is standardized to make geographic and business focus variables comparable and results easier to interpret:

$$Bus\_focus = \frac{HHI_{BF} - 1/4}{1 - 1/4} \qquad (6)$$

This ratio takes the values from 0 to 1.

At this point it is worth pointing out that business diversification is included in the analysis as a control variable because Bodnar et al. (1998) show that by not including both types of diversification together into the model, there emerges an omitted variable problem.

4.3 Control variables

Similarly as in other similar studies, the following control variables are used.

Leverage (Book_lev)
Book Leverage = book value of total liabilities/book value of equity and total liabilities

A positive relationship between leverage and implied cost of equity capital is expected to be found (Modigliani and Miller, 1958; Dhaliwal et al., 2006; Attig et al., 2008). The reason for using a book and not market leverage is because in the model are also included market anomaly variables; therefore the problem of correlation between independent variables needs to be avoided.

Credit risk (Llprov_loans)
Loan portfolio risk = Loan loss provisions/Loans

This paper follows other authors who use this variable (Acharya et al., 2006; Iannotta et al. 2007) as a proxy for credit risk, while the intuition is the following. Most probably investors do not know the true quality of bank loan portfolios, but they can obtain some information on their quality on the basis of loan loss provisions made by the bank. It can be expected that investors perceive the amount of loan loss provisions that the management reserves to cover unexpected future losses on loan defaults as a signal of banks' credit risk. This signal will affect the stock market assessment of the bank's risk. Consequently, a positive relationship between this variable and implied cost of equity capital is expected to be found.

Market risk (Volatility)
Volatility = the standard deviation of monthly prices over the last 12 months divided by the average of the monthly price over the last 12 months, calculated in June of every year.

There are several empirical studies (for instance: Attig et al., 2008, Boubakri et al., 2008) who choose to capture firm's market risk with this variable. Moreover, these studies find a positive and statistically significant relationship between volatility and cost of equity capital. And this is also the expected sign of this variable in this study.

Furthermore, the choice to prefer this variable over beta to proxy for banks' market risk is because recent empirical literature finds that beta exhibits little or no association with the implied cost of equity capital (see for instance, Gebhardt et al., 2001).

The dispersion of analysts' earnings forecasts (Ln_kv)
The dispersion of analysts' earnings forecasts = natural logarithm of the coefficient of variation of analysts' one-year-ahead earnings per share forecasts as reported by I/B/E/S in June of every year.

Given that analyst forecast dispersion should be positively related to information asymmetry, there should be a positive relationship also between this variable and implied cost of equity capital (Boubakri et al., 2008; Dhaliwal et al., 2006).

Book to market (Bm)
Book to market = book value of equity capital/market value of equity capital.

A positive relationship between book to market and implied cost of equity capital is expected to be found (Fama and French, 1992; Attig et al., 2008; Chen et al., 2009).

Price momentum (Momentum)
Price Momentum = a buy and hold return on the bank's stock over the period: beginning of June (t-1) until the end of May (t).

There are theories supporting a positive and a negative relationship between this variable and implied cost of equity capital. On one hand, Jegadeesh and Titman (1993) argue that price momentum is a risk proxy, so it can be expected to find that high momentum stocks also have higher implied risk premium and that there should be a positive relationship between momentum variable and cost of equity capital. While on the other hand, Guay et al. (2005) suggest including recent stock returns as a control variable for sluggishness in analyst forecasts. This bias occurs because revisions in investors' expectations about future earnings are immediately reflected in stock price, but they might not be also incorporated into the analyst forecasts immediately. This implies that there might be a negative correlation between momentum and cost of equity capital. Similar approach and finding of a negative relationship between momentum and implied cost of equity capital is also found in Chen et al. (2009).

Year dummy
Year dummies are included as control variables for common trends or business cycle effects.

Furthermore, even the following variables were considered for inclusion in the model: ROAA, cost-to-income ratio, natural logarithm of total assets, loans over total assets, deposits over total assets and state dummies, but given that they are not statistically significant and that including them in the model does not lead to different results they are not included in the analysis.

4.4 Estimation methodology

4.4.1 Descriptive statistics and correlation between variables

The sample includes 93 banks over the sample period 1995-2010. The sample includes two types of banks according to specialization: commercial banks and bank holding companies, with most banks in the sample being bank holding companies.

From Table 2 it can be seen that banks have higher business diversification than the geographic one. The standardized Herfindahl-Hirschman index of business focus is for all bank years on average 0.398 whilst the standardized Herfindahl-Hirschman index of geographic focus is for all bank years on average 0.622. The variability of the geographic focus is higher than the variability of business focus. The standard deviation of the standardized Herfindahl-Hirschman index of geographic focus is for all bank years 0.311, while the standard deviation of the standardized Herfindahl-Hirschman index of business focus is for all bank years 0.183. As expected, banks have high leverage - leverage is for all bank years on average 0.910. The highest correlation coefficient between different pairs of independent variables included in the model is between volatility and kv (0.496) (see Table 3).

Table 4 reports descriptive statistics for various models of implied cost of equity capital. As expected, there are differences between various models. The highest cost of equity capital for all bank years is calculated by the OJ model (0.092) whereas the lowest is calculated by the GLS model (0.080). In Table 5 correlation coefficients between all pairs of different models of implied cost of equity capital are reported. As is seen, they are all statistically significant. The highest correlation coefficient is between OJ and Easton model (0.997) while the lowest correlation coefficient is between OJ and GLS model (0.663).

The overall average cost of equity capital is 8.5% (see Table 6). Moreover, the average increase in cost of equity capital from one year to another was the highest in the period 2007-2008, when it amounted to 4.40 percentage points. The overall average of geographic focus for US banks is 0.622. The average of geographic focus variable was increasing over the sample period.

4.4.2 Estimation techniques applied

Dynamic panel data: Blundell-Bond estimator

This paper begins investigating the relationship between geographic focus and banks' cost of equity capital by including into the fixed effects model one lag of the dependent variable as a regressor and estimates the model by using generalized method of moments (GMM). The main theoretical reason for using dynamic panel data is because it is modeling a partial adjustment based approach and the two commonly used dynamic panel estimators. Arellano-Bond (1991) and Blundell-Bond (1998). Arellano-Bond estimator starts by differencing all

regressors and uses the GMM. Its potential weakness is that the lagged levels are often poor instruments for first-differenced variables, and this is especially so if the variables are close to a random walk. Blundell and Bond (1998) use an additional assumption: first differences of instrumenting variables are uncorrelated with fixed effects, which allows the introduction of more instruments and can improve efficiency. It builds a system of two equations, the original equation and the transformed one. Therefore, it can be said that Blundell and Bond (1998) improve the properties of the standard first-differenced GMM estimator by using additional moment conditions to obtain an estimator with improved precision and better finite-sample properties. This is the reason why this one is applied in the analysis.

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \sum_{n=2}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (7)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus, as described in section 4.2.1;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3. (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

1-step and 2-step[9] GMM estimation results are presented for the main model and 2-step GMM estimation results for additional questions addressed. All the estimations are performed with the program xtabond2 in Stata (Roodman, 2006). Robust standard errors are used. For 1-step estimation, robust specifies that the standard errors are consistent in the presence of heteroskedasticity and autocorrelation within panels. While in the 2-step estimation where the errors are already robust but usually downward bias, the Windmeijer's finite-sample correction for the two-step covariance matrix is calculated to correct for this. In the case where the Blundell and Bond estimator is used, two assumptions need to be tested. The one that the disturbances are serially uncorrelated (no second-order serial correlation in the first-differenced residuals) and that the instruments are valid instruments; that they are uncorrelated with the first-differenced residuals (Hansen test of over-identifying restrictions).

Heckman two-stage estimation technique

To check the robustness of the results the Heckman's (1979) two-stage estimation technique is applied next. Heckman's (1979) two-stage estimation technique allows to control for self-

[9] Two-step estimation means that the two-step estimator is calculated instead of the one-step one, that the covariance matrix is estimated by using the first-step residuals.

selection of firms that diversify, for the endogeneity of the firms' diversification decision. An approach similar to the one used in Campa and Kedia (2002) is used. The main findings of Campa and Kedia (2002), who analyze business diversification of US nonfinancial firms over the period 1978-1996, are that firms self-select into becoming diversified and that self-selection explains the diversification discount.

A probit regression, with a dummy variable whether the bank is diversified as the dependent variable, is estimated in the first-step, while the choice of explanatory variables is based on variables found to influence the firms' decision to diversify in Campa and Kedia (2002).

These variables are:

Industry instruments

1. Average propensity to diversify abroad = the fraction of all banks in a state, which have their operation diversified in other states. If the bank reports deposit diversification in at least one other state it is defined as diversified. The analysis includes public and private banks with total assets larger than 5 billion US dollars. (Industry_div).
   It can be expected that the higher the fraction of geographically diversified banks in other states, the more attractive other states are for banks to diversify their operations there.

Time trends

2. Number of completed M&A transactions (financial sector only) for each state in every year over the average number of completed M&A transactions (financial sector only) for each state over the sample period (N_deals). Database Zephyr is used to obtain this information, but because this one does not report data for the time period before 1997 the analysis is limited to the time period 1997-2010.
   Campa and Kedia (2002) suggest capturing time trends by the existence of merger waves. It is intuitive that the more active the market for M&A is, the higher the probability that a firm diversifies.

Firm specific instruments

3. Total assets over average of total assets for each state in every year (in order to calculate state average listed and private banks from a state with total assets larger than 5 billion US dollars are considered). (Ta)
4. A dummy variable whether the bank is listed on NYSE, AMEX or NASDAQ. (Se)
5. A dummy variable whether the firm belongs to the S&P 500 stock index. (Sp500)
   Campa and Kedia (2002) argue that firms are more likely to diversify if they are listed on the major exchanges (NYSE, AMEX or NASDAQ). This is so because those firms that are listed on these exchanges have a higher visibility, lower information asymmetry (they are followed by more financial analysts) and higher liquidity. According to Campa and Kedia (2002), whether a firm is included in S&P index, is a control variable for liquidity

It is important to stress that even if the two variables, total assets and numbers of M&A transactions in a given year are standardized, results remain consistent if variables are not standardized and are used as Campa and Kedia (2002) suggest.

In the second stage, the cost of equity capital is regressed on the standardized Herfindahl-Hirschman index of geographic focus, other independent variables and the self-selection parameter (lambda).

The system can be descried as:

$$D_{it} = 1 \quad if \quad Z_i\gamma + \eta_i > 0 \tag{8}$$

$$D_{it} = 0 \quad if \quad Z_i\gamma + \eta_i \leq 0 \tag{9}$$

$$k_{avergae\ it} = d_0 + d_1 X_{it} + d_2 Geog\_focus_{it-1} + d_\lambda \lambda + \varepsilon_{it} \tag{10}$$

D_{it} … a diversification dummy equal to 1 if the firm operates in more than one geographic segment;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus, as described in section 4.2.1;

Z_i … a set of explanatory variables described above;

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1.1;

X_{it} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3. (business focus, banks' risk, market anomaly variables, information asymmetry variable);

ε_{it} … an error term;

Under the assumption that the error terms are bivariate normal, the system can be estimated as Heckman selection model.

Results

Main results of the two estimation techniques are similar and can be summed up as:

- Cost of equity capital is influenced by the cost of equity capital of the previous year.
- The regression coefficient of geographic focus is negative and statistically significant. It can be interpreted by saying: if the standardized Herfindahl-Hirschman index of geographic focus increases by one standard deviation, other things being equal, then expected cost of equity capital decreases by 19 basis points.
- Regression coefficient of business focus variable is not statistically significant.

- Among the three measures of risk only the regression coefficient of market risk (volatility) is always positive and statistically significant. Regression coefficients of leverage and loan loss provisions over loans are mostly insignificant.
- As expected, regression coefficient of the coefficient of variation is positive, while regression coefficient of momentum is negative. Regression coefficients of the book-to-market ratio is mostly insignificant.
- There is self selection, but it does not affect the main result.

4.5 Additional questions addressed

Results of this section analysis are reported in Table 9.

Substituting the implied cost of equity capital with inverse price-earnings ratio as a proxy for cost of equity capital

To check the results further, an alternative model of cost of equity capital, an inverse of price-earnings ratio, is also considered. If it is assumed that next year's earnings forecast is sufficient for valuation, cost of equity capital can be estimated as the inverse of the forward PE ratio.

$$k_{EP} = \frac{FEPS_{t+1}}{P_t} \tag{14}$$

P_t … market price of a bank's common stock in June of year t;
$FEPS_{t+1}$ … I/B/E/S consensus earnings per share forecast for the next year at time t (forecasts are collected as of June of every year).

As can be seen in Table 9 the main results do not differ much from the ones in the main section of the results.

Various business streams and the average cost of equity capital

It is intuitive that the regression coefficients of different revenue streams might have a different sign. Therefore, the following model is estimated in order to check this intuition:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Int_{it-1} + \alpha_3 Com_{it-1} + \alpha_4 Trad_{it-1} + \\ + \sum_{n=5}^{N} \alpha_n X_{nit} + c_i + u_{it} \tag{15}$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus, as described in section 4.2.1;

Int_{it-1} … net interest revenue in total operating revenue,

Com_{it-1} … net commission and fee revenue in total operating revenue;

$Trad_{it-1}$ … net trading revenue in total operating revenue;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3 (banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

As seen in Table 9, regression coefficients of different revenue streams do not have a different sign. Moreover, these variables are not statistically significant even if net interest revenue over total operating revenue, net commission and fee revenue over total operating revenue and net trading revenue over total operating revenue are included in the analysis individually – one at a time.

Is the relationship nonlinear?

Investors might find a certain degree of diversification beneficial. Therefore the relationship between geographic diversification and implied cost of equity capital might be nonlinear. In order to check this possibility, the following models are estimated.

A quadratic function:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Geog\_focus^2{}_{it-1} + \sum_{n=3}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (16)$$

A cubic function:

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \alpha_2 Geog\_focus^2{}_{it-1} + \alpha_3 Geog\_focus^3{}_{it-1} + \\ + \sum_{n=4}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (17)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus, as described in section 4.2.1;

$Geog\_focus^2{}_{it-1}$ … a square of the standardized Herfindahl-Hirschman index of geographic focus;

$Geog\_focus^3_{it-1}$ … a cube of the standardized Herfindahl-Hirschman index of geographic focus;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

From Table 9 it can be seen that neither a quadratic function nor a cubic function are appropriate to explain the relationship addressed in this analysis.

Substituting leverage variable with Tier 1 or Z-score ratio

The model is estimated also by substituting leverage with Tier 1 and Z-score. As seen in Table 9, in this case the results do not change much, but the number of observations drops.

$$k_{average_{it}} = \gamma_1 k_{average_{i,t-1}} + \alpha_1 Geog\_focus_{it-1} + \sum_{n=2}^{N} \alpha_n X_{nit} + c_i + u_{it} \quad (18)$$

$k_{average_{it}}$ … the average cost of equity capital, as described in section 4.1.1;

$Geog\_focus_{it-1}$ … the standardized Herfindahl-Hirschman index of geographic focus, as described in section 4.2.1;

X_{nit} … control variables include: year dummies and other variables commonly used in similar studies of cost of equity capital, as described in section 4.3 (business focus, banks' risk, market anomaly variables, information asymmetry variable);

c_i … individual-specific effects;

u_{it} … an error term.

5 Conclusion

Interstate diversification of American banks was initially constrained by severe regulation; however, over the years this regulation was weakened and the barriers, which disable banks to expand geographically, were reduced. As a result we have witnessed increasing trend in geographic diversification of American banks in the past decade. In relation to this fact, the question of whether interstate geographic diversification has a statistically significant effect on banks' cost of equity capital has been left unanswered. Therefore, this study addresses this issue. It investigates the link between the banks' interstate diversification and their cost of equity capital in order to try to find some further evidence to answer the question of whether banks should diversify geographically or stay focused to achieve lower cost of equity capital. It is also important to add that the reason why the issue of diversification is an important one

is because it is related to the optimum degree of diversification, while the cost of equity capital is an important variable because it is the determinant of shareholders' value.

Due to the fact that from a theoretical point of view the effects of geographic diversification on cost of equity capital are controversial, this study tries to contribute to this debate by examining the relationship between the banks' geographic diversification and their cost of equity capital. It tries to determine if positive or negative net effects prevail.

In the empirical analysis various estimation techniques are used. The sample consisted of the largest public banks over the period 1995-2010. The implied cost of equity capital is estimated by implying four commonly used models (Claus and Thomas, 2001; Easton, 2004; Gebhardt et al., 2001; Ohlson and Juettner-Nauroth, 2005). To measure diversification between major geographic areas in which the bank operates, a deposit based Herfindahl-Hirschman Index is constructed.

The main finding of this analysis is that there is a positive relationship between banks' geographic diversification and their cost of equity capital. This finding is consistent with the agency theory, internal capital market and investors' negative reaction to this banks' business strategy. Furthermore, empirical studies, which find a negative net effect of geographic diversification, explain it by pointing out that banks in foreign markets develop more risky credit portfolios or business practices, that the poor monitoring incentives may arise in these new markets, or by suggesting that managers are not skilled enough for this business strategy. Given that the main finding is similar as in the European sample (though the effect is smaller in this sample than it is in the European one) even the policy implication is similar, and it is that regulators should think carefully before imposing regulation which would incentivize banks to increase the level of their interstate diversification. This is so because apparently investors view this banks' business strategy as either value reducing or they view such banks as more risky. In either case these banks appear to have a higher cost of equity capital, which is against the goals of regulators and supervisors.

References

Acharya, V.V., Hasan, I. and Saunders, A. (2006) Should banks be diversified? Evidence from individual bank loan portfolios, *Journal of Business* **79**, 1355–1412.

Adams, R. and Mehran, H. (2003) Is corporate governance different for bank holding companies?, *Economic Policy Review* **9**, 123–142.

Agmon, T. and Lessard, D.R. (1977) Investor recognition of corporate international diversification, *Journal of Finance* **32**, 1049–1055.

Akhigbe, A. and Whyte, A.M. (2003) Changes in market assessment of bank risk following the Riegle-Neal Act of 1994, *Journal of Banking and Finance* **27**, 87–102.

Amihud, Y. and Lev, B. (1981) Risk reduction as managerial motive for conglomerate mergers, *Bell Journal of Economics* **12**, 605– 617.

Arellano, M. & Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**, 277–297.
Attig, N., Guedhami, O. and Mishra, D. (2008) Multiple large shareholders, control contest, and implied cost of equity, *Journal of Corporate Finance* **14**, 721–737.
Berger, A.N., DeYoung, R., Genay, H. and Udell, G.F. (2000) Globalization of financial institutions: Evidence from cross-border banking performance, *Brookings-Wharton Papers on Financial Services* **3**, 23–120.
Berger, A.N. and DeYoung, R. (2001) The effects of geographic expansion on bank efficiency, *Journal of Financial Services Research* **19**, 163–184.
Berger, A.N. and Hannan, T.H. (1998) The efficiency cost of market power in the banking industry: A test of the `quiet life' and related hypotheses, *Review of Economics and Statistics* **80**, 454–465.
Bliss, R. and Rosen, R. (2001) CEO compensation and bank mergers, *Journal of Financial Economics* **61**, 107–138.
Blundell, R. and Bond, S. (1998) Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* **87**, 115–143.
Bodnar, G.M., Tang, C. and Weintrop, J. (1998), Both sides of corporate diversification: The value impacts of geographic and industrial diversification, unpublished working paper, Wharton School, University of Pennsylvania.
Boyd, J. and Prescott, E. (1986) Financial intermediary coalitions, *Journal of Economic Theory* **38**, 211–232.
Botosan, C.A. and Plumlee, M.A. (2005) Assessing alternative proxies for the expected risk premium, *The Accounting Review* **80**, 21–53.
Boubakri, N., Guedhami, O., Mishra, D. and Saffar, W. (2008), Political connections and the cost of equity capital, unpublished working paper, HEC Montreal.
Burgman, T.A. (1996) An empirical examination of multinational corporate capital structure, *Journal of International Business Studies* **27**, 553–570.
Campa, J.M. and Kedia, S. (2002) Explaining the diversification discount, *Journal of Banking and Finance* **57**, 1731–1762.
Caves, R.E. (1971) International corporations: The industrial economics of foreign investment, *Econometrica* **38**, 1–27.
Chen, K.C.W., Chen, Z.H. and Wei, K.C.J. (2009) Legal protection of investors, corporate governance, and the cost of equity capital, *Journal of Corporate Finance* **15**, 273–289.
Chen, K.C.W., Chen, ZH. and Wei, K.C.J. (2011) Agency Costs of Free Cash Flow and the Effect of Shareholder Rights on the Implied Cost of Equity Capital, *Journal of Financial and Quantitative Analysis* **46**, 171-207.
Chong, B.S. (1991) The effects of interstate banking on commercial banks' risk and profitability, *Review of Economics and Statistics* **73**, 78–84.

Claus, J. and Thomas, J. (2001) Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* **56**, 1629–1666.
Cochrane, J.H. (2010), Discount Rates, unpublished working paper, Booth School of Business, University of Chicago.
Consensus Economics, *Consensus forecasts: A digest of international economic forecasts.*
De Haas, R. and Van Lelyveld, I. (2010) Internal capital markets and lending by multinational bank subsidiaries, *Journal of Financial Intermediation* **19**, 1–25.
DeLong, G. (2001) Stockholder gains from focusing versus diversifying bank mergers, *Journal of Financial Economics* **59**, 221–252.
Deng, S.E. and Elyasiani, E. (2008) Geographic diversification, bank holding company value, and risk, *Journal of Money, Credit and Banking* **40**, 1217–1238.
Denis, D., Denis, D. and Sarin, A. (1997) Agency problems, equity ownership, and corporate diversification, *Journal of Finance* **52**, 135–160.
Dhaliwal, D., Heitzman, S. and Li, O.Z. (2006) Taxes, leverage, and the cost of equity capital, *Journal of Accounting Research* **44**, 691–723.
Diamond, D. (1984) Financial intermediation and delegated monitoring, *Review of Economic Studies* **51**, 393–414.
Easton, P. (2004) PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital, *Accounting Review* **79**, 73–95.
Easton, P.D. and Monahan, S.J. (2005) An evaluation of accounting-based measures of expected returns, *Accounting Review* **80**, 501–538.
Easton, P. and Sommers, G. (2007) Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts, *Journal of Accounting Research* **45**, 983–1015.
El Ghoul, S., Guedhami, O. and Pittman, J. (2011) The role of IRS monitoring in equity pricing in public firms, *Contemporary Accounting Research* **28**, 643-674.
Elsas, R., Hackethal, A. and Holzhäuser, M. (2006), The anatomy of bank diversification, unpublished working paper, University of Munich.
Elton, E.J. (1999) Expected return, realized return, and asset pricing tests, *Journal of Finance* **54**, 1199–1220.
Fama, E. and French, K. (1992) The cross sections of expected stock returns, *Journal of Finance* **47**, 427–466.
Fama, E. and French, K. (1997) Industry costs of equity, *Journal of Financial Economics* **43**, 153–193.
Fatemi, A.M. (1984) Shareholder benefits from corporate international diversification, *Journal of Finance* **39**, 1325–1344.
Francis, J., Lafond, R., Olsson, P.M. and Schipper K. (2005) The market pricing of accruals quality, *Journal of Accounting and Economics* **39**, 295-327.
Frankel, R. and Lee, C. (1998) Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and Economics* **25**, 283–319.

Garcia-Herrero, A. and Vazquez, F. (2007), International diversification gains and home bias in banking, unpublished working paper, International Monetary Fund.
Gebhardt, W., Lee, C. and Swaminathan, B. (2001) Towards an implied cost of capital, *Journal of Accounting Research* **39**, 135–176.
Gode, D. and Mohanram, P. (2003) Inferring the cost of capital using the Ohlson-Juettner Model, *Review of Accounting Studies* **8**, 399–431.
Guay, W.R., Kothari, S.P. and Shu, S. (2005), Properties of implied cost of capital using analysts' forecasts, unpublished working paper, Massachusetts Institute of Technology.
Guedhami, O. and Mishra, D. (2009) Excess control, corporate governance, and implied cost of equity: International evidence, *Financial Review* **44**, 489-524.
Hail, L. and Leuz, C. (2006) International differences in the cost of equity capital: Do legal institutions and securities regulation matter?, *Journal of Accounting Research* **44**, 485–531.
Hail, L. and Leuz, C. (2009) Cost of capital effects and changes in growth expectations around U.S. cross-listings, *Journal of Financial Economics* **93**, 428-454.
Hann R., Ogneva, M. and Ozbas, O. (2009), Corporate diversification and the cost of capital, unpublished working paper, Stanford University.
Hayden, E., Porath, D. and Westernhagen, N. (2007) Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios, *Journal of Financial Services Research* **32**, 123–140.
Heckman, J. (1979) Sample selection bias as specification error, *Econometrica* **47**, 153–161.
Hribar, P. and Jenkins, N. (2004) The effect of accounting restatements on earnings revisions and the estimated cost of capital, *Review of Accounting Studies* **9**, 337-356.
Hughes, J.P., Lang, W.W., Mester, L.J. and Moon, C.G. (1999) The dollars and sense of bank consolidation, *Journal of Banking and Finance* **23**, 291–324.
Iannotta, G., Nocera, G. and Sironi, A. (2007) Ownership structure, risk, and performance in the European banking industry, *Journal of Banking and Finance* **31**, 2127–2149.
Jegadeesh, N. and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* **48**, 65–91.
Jensen, M.C. (1986) Agency costs of free cash flow, corporate finance and takeovers, *American Economic Review* **76**, 323–329.
Jensen, M.C. (1993) The modern industrial revolution, exit, and the failure of internal control systems, *Journal of Finance* **48**, 831–880.
King, M.R. (2009), The cost of equity for global banks: A CAPM perspective from 1990 to 2009, unpublished working paper, Bank for International Settlements.
Kogut, B. (1983) *Foreign Direct Investment as a Sequential Process,* The MIT Press, Cambridge, Massachusetts.
Lee, K.C. and Kwok, C.C.Y. (1988) Multinational corporations vs. domestic corporations: International environmental factors and determinants of capital structure, *Journal of International Business Studies* **19**, 195–217.
Lewellen, W. (1971) A pure financial rationale for the conglomerate merger, *Journal of Finance* **26**, 521–537.

Maccario, A., Sironi, A. and Zazzara, C. (2002), Is banks' cost of equity capital different across countries? Evidence from the G10 countries major banks, unpublished working paper, Libera Università Internazionale degli Studi Sociali Guido Carli.
Modigliani, F. and Miller, M.H. (1958) The cost of capital, corporation finance, and the theory of investment, *American Economic Review* **48**, 261–297.
Morck, R. and Yeung, B. (1991) Why investors value multinationality, *Journal of Business* **64**, 165–186.
Morgan, D.P. (2002) Rating banks: Risk and uncertainty in an opaque industry, *American Economic Review* **92**, 874–888.
Morgan, D.P. and Samolyk, K. (2003), Geographic diversification in banking and its implications for bank portfolio choice and performance, unpublished working paper, Bank for International Settlements.
Ohlson, J. and Juettner-Nauroth, B. (2005) Expected EPS and EPS growth as determinants of value, *Review of Accounting Studies* **10**, 349–365.
Pastor, L., Sinha, M. and Swaminathan, B. (2008) Estimating the intertemporal risk-return tradeoff using the implied cost of capital, *Journal of Finance* **63**, 2859–2897.
Rajan, R., Servaes, H. and Zingales, L. (2000) The cost of diversity: The diversification discount and inefficient investment, *Journal of Finance* **55**, 35−80.
Roodman, D. (2009) How to do xtabond2: An introduction to difference and system GMM in Stata, *The Stata Journal* **9**, 86–136.
Rowland, P.F. and Tesar, L.L. (2004) Multinationals and the gains from international diversification, *Review of Economic Dynamics* **7**, 789–826.
Scharfstein, D.S. and Stein, J.C. (2000) The dark side of internal capital markets: Divisional rent-seeking and inefficient investment, *Journal of Finance* **55**, 2537−2564.
Schmautzer, D. (2006), Cross-border bank mergers: Who gains and why?, unpublished working paper, University of Muenster.
Schmid, M.M. and Walter, I. (2009) Do financial conglomerates create or destroy economic value?, *Journal of Financial Intermediation* **18**, 193–216.
Stein, J.C. (1997) Internal capital markets and the competition for corporate resources, *Journal of Finance* **52**, 111–133.
Stiroh, K.J. (2004) Do community banks benefit from diversification?, *Journal of Financial Services Research* **25**, 135–160.
Stiroh, K.J. and Rumble, A. (2006) The darkside of diversification: The case of U.S. financial holding companies, *Journal of Banking and Finance* **30**, 2131–2161.
Waldron, D.G. (2006) International diversification and the cost of capital: Is more necessarily better?, *International Business & Economics Research Journal* **5**, 23-38.
Wulf, J. (1999), Influence and inefficiency in the internal capital market: Theory and evidence, unpublished working paper, University of Pennsylvania.
Zimmer, S.A. and McCauley, R.N. (1991) Bank cost of capital and international competition, *Federal reserve bank of New York quarterly review*, 33–59.

Tables

Table 1: Banks' cost of equity capital estimated by researchers

	Zimmer & McCauley (1991)	Maccario et al. (2002)	King (2009)	
	1984-1990	1993-2001	1993-2001	2002-2009
France		7.7%	10.6%	7.3%
Germany	6.9%	7.0%	11.4%	9.0%
United Kingdom	9.8%	8.9%	9.5%	6.6%
Japan	3.1%	2.8%	12.0%	11.2%
Canada	10.3%	12.0%	10.7%	5.4%
United States	11.9%	8.8%	10.4%	7.2%
Switzerland	5.3%	8.2%		
Belgium		8.9%		
Spain		8.0%		
Italy		7.6%		
Netherland		9.0%		
Sweden		9.7%		

Source: King (2009), Maccario et al. (2002), Zimmer & McCauley (1991).

Table 2: Descriptive statistics for independent variables

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
Geog_focus	0.599	0.622	0.311	0.340	0.973	1025
Bus_focus	0.377	0.398	0.183	0.258	0.506	1025
Leverage	0.914	0.910	0.025	0.900	0.926	1025
Llprov_loans	0.003	0.006	0.013	0.002	0.005	1025
Ln_kv	-3.951	-3.749	0.975	-4.357	-3.384	1025
Momentum	0.093	0.122	0.316	-0.061	0.288	1025
Bm	0.492	0.745	1.497	0.367	0.643	1025
Volatility	0.086	0.104	0.077	0.059	0.125	1025
Total_assets	16,900,000	88,800,000	249,000,000	9,134,700	52,100,000	1025

Notes: Total assets are in 1000 US dollars. Geog_focus is geographic focus; its calculation is described in section 4.2.1. Bus_focus is business focus; its calculation is described in section 4.2.2. Leverage is the ratio between the book value of total liabilities and sum of book value of equity and total liabilities. Llprov_loans is the ratio between loan loss provisions and loans. Ln_kv is a natural logarithm of the coefficient of variation of analysts' one-year-ahead earnings per share forecasts as reported by I/B/E/S. Momentum is a buy and hold return on the bank's stock over the period: beginning of June (t-1) until the end of May (t). Bm is the ratio of book value of equity capital to market value of equity capital. Volatility is the standard deviation of monthly prices over the last 12 moths divided by the average of the monthly price over the last 12 moths.

Table 3: Correlation coefficients between independent variables

Variable	Geog_ focus	Bus_ focus	Leverage	Llprov_ loans	Ln_kv	Momentum	Bm	Volatility
Bus_focus	0.336***	1						
Leverage	-0.033	-0.141***	1					
Llprov_ loans	0.012	-0.200***	-0.184***	1				
Ln_kv	0.080***	0.044	-0.112***	0.264***	1			
Momentum	0.164***	0.086***	0.049	0.019	-0.204***	1		
Bm	-0.120***	-0.033	-0.066**	0.070**	0.146***	-0.043	1	
Volatility	0.119***	0.034	0.058*	0.351***	0.496***	-0.059	0.094***	1
Ln_total_ assets	-0.547***	-0.515***	0.091***	0.147***	0.044	-0.186***	0.078***	-0.007

Notes: Variables are defined in Table 2.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 4: Descriptive statistics for implied cost of equity capital models

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
CT	0.073	0.081	0.033	0.064	0.088	1025
GLS	0.078	0.080	0.021	0.067	0.088	1025
OJ	0.083	0.092	0.037	0.073	0.100	1025
Easton	0.079	0.088	0.036	0.068	0.095	1025
Average	0.078	0.085	0.029	0.069	0.092	1025

Notes: CT is the cost of equity capital, estimated by the Claus and Thomas (2001) model. GLS is the cost of equity capital estimated by Gebhardt et al. (2001) model. OJ is the cost of equity capital estimated by Ohlson and Juettner Nauroth (2005) model. Easton is cost of equity capital estimated by Easton (2004) model. The calculation of variables is defined in Section 4.1. Average the average cost of equity capital, calculated as the average of these four models.

Table 5: Correlation coefficients between different models of implied cost of equity capital

	CT	GLS	OJ	Easton
CT	1			
GLS	0.710***	1		
OJ	0.763***	0.663***	1	
Easton	0.776***	0.701***	0.997***	1
Average	0.890***	0.808***	0.960***	0.970***

Notes: Variables are defined in Table 4.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 6: Mean implied cost of equity capital and geographic focus (diversification) by year

Year	N	Mean of Average	Mean of Geog_focus
1995	54	0.085	0.704
1996	57	0.086	0.707
1997	58	0.071	0.698
1998	65	0.066	0.685
1999	70	0.074	0.663
2000	72	0.096	0.641
2001	76	0.076	0.629
2002	76	0.080	0.607
2003	81	0.086	0.595
2004	79	0.080	0.596
2005	75	0.078	0.591
2006	73	0.076	0.584
2007	68	0.079	0.583
2008	52	0.123	0.534
2009	33	0.138	0.579
2010	36	0.126	0.544
1995-2010	1025	0.085	0.622

Table 7: Estimated dynamic panel data model (Blundell-Bond dynamic panel data estimator): geographic focus and banks' cost of equity capital

		DPD 1 step		DPD 2 step	
Variables		**(1)**	**(2)**	**(3)**	**(4)**
L.Average	**b**	**0.313*****	**0.322*****	**0.316*****	**0.313*****
	se	0.067	0.065	0.068	0.062
Geog_focus	**b**	**-0.007*****	**-0.007*****	**-0.006***	**-0.006***
	se	0.003	0.003	0.003	0.003
Bus_focus	**b**	**0.002**	**0.001**	**0.000**	**-0.001**
	se	0.005	0.005	0.006	0.005
Leverage	**b**	**-0.003**	**-0.005**	**0.000**	**0.005**
	se	0.032	0.032	0.037	0.042
Llprov_loans	**b**	**0.005**	**-0.003**	**-0.021**	**-0.012**
	se	0.075	0.073	0.078	0.075
Ln_kv	**b**	**0.007*****	**0.006*****	**0.007*****	**0.006*****
	se	0.001	0.001	0.001	0.001
Momentum	**b**	**-0.036*****	**-0.036*****	**-0.036*****	**-0.037*****
	se	0.003	0.003	0.004	0.004
Bm	**b**	**0.001***	**0.001***	**0.001**	**0.001**
	se	0.001	0.001	0.001	0.001
Volatility	**b**	**0.144*****	**0.143*****	**0.144*****	**0.143*****
	se	0.020	0.019	0.021	0.022
	constant	**0.092*****	**0.093*****	**0.088*****	**0.086****
	se	0.028	0.028	0.033	0.038
	N	954	954	954	954
	Number of groups	93	93	93	93
	Number of instruments	142	87	142	87
	Wald chi2	1119.48	1104.01	1015.88	1016.54
	z(AR1)	-4.48	-4.53	-3.62	-3.64
	z(AR2)	-0.54	-0.56	-0.56	-0.50
	Hansen J test	79.1	71.94	79.1	71.94

Notes: Variables are defined in Table 2. Dependent variable is the cost of equity capital. b refers to the regression coefficient, se to robust standard error. Instruments: all lags of Average as the GMM instruments and year dummies and explanatory variables as IV instruments (142 instruments) and lags of Average 1 to 4 as the GMM instruments and year dummies and explanatory variables as IV instruments (87 instruments).
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 8: Estimated Heckman selection model: geographic focus and banks' cost of equity capital

Variables		(1)
Geog_focus	**b**	**-0.007*** **
	se	0.003
Bus_focus	**b**	**0.005**
	se	0.005
Leverage	**b**	**-0.052**
	se	0.038
Llprov_loans	**b**	**0.666*** **
	se	0.229
Ln_kv	**b**	**0.006*** **
	se	0.001
Momentum	**b**	**-0.032*** **
	se	0.005
Bm	**b**	**0.001***
	se	0.001
Volatility	**b**	**0.190*** **
	se	0.032
	constant	**0.133*** **
	se	0.035
Ta	**b**	**0.323*** **
	se	0.090
Industry_div	**b**	**3.047*** **
	se	0.218
Se	**b**	**-0.299**
	se	5.219
Sp500	**b**	**-0.234**
	se	0.148
N_deals	**b**	**0.064**
	se	0.082
	constant	**-0.559**
		5.218
	Lambda	-0.005**
	N	915
	Censored N	194

Notes: Variables are defined in Table 2. In addition, Ta is total assets over the state average of total assets, Industry_div is the fraction of all banks in the state, which have their operation diversified in other states, Se is a dummy variable whether the banks it is listed on NYSE, AMEX or NASDAQ stock exchange, Sp500 is a dummy variable whether the banks belongs to the S&P 500 stock index, N_deals is a number of completed M&A transactions in a given year over the average number of M&A transactions in each state over the sample period. b refers to the regression coefficient, se standard error estimated by Jackknife option. ***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 9: Additional questions addressed

Variables		(1)	(2)	(3)	(4)	(5)	(6)	(7)
L.Average	**b**		**0.285*** **	**0.320*** **	**0.322*** **	**0.399*** **	**0.305*** **	**0.322*** **
	se		0.061	0.060	0.060	0.071	0.063	0.062
L.ep	**b**	**0.517*** **						
	se	0.098						
Geog_focus	**b**	**-0.003***	**-0.007** **	**-0.024**	**-0.015**	**-0.005***	**-0.005***	**-0.005***
	se	0.002	0.003	0.016	0.040	0.003	0.003	0.003
Int	**b**		**-0.011**					
	se		0.009					
Com	**b**		**-0.014**					
	se		0.012					
Trad	**b**		**-0.055**					
	se		0.053					
Geog_focus^2	**b**			**0.016**	**-0.002**			
	se			0.014	0.078			
Geog_focus^3	**b**				**0.010**			
	se				0.047			
Tier 1	**b**					**-0.031**		
	se					0.048		
Zscore$_{mv}$	**b**						**-0.004*** **	
	se						0.001	
Zscore$_{bv}$	**b**							**-0.003*** **
	se							0.001
Bus_focus	**b**	**0.007***		**-0.001**	**-0.001**	**0.001**	**-0.002**	**-0.001**
	se	0.004		0.005	0.005	0.006	0.005	0.005
Leverage	**b**	**0.076***	**0.018**	**0.001**	**0.008**			
	se	0.043	0.038	0.041	0.041			

Cont. of Table 9

Llprov_loans	**b**	**0.124****	**-0.006**	**-0.024**	**-0.023**	**0.199**	**-0.039**	**-0.043**
	se	0.052	0.072	0.072	0.077	0.189	0.064	0.063
Ln_kv	**b**	**-0.003**	**0.006*****	**0.006*****	**0.006*****	**0.005*****	**0.005*****	**0.005*****
	se	0.002	0.001	0.001	0.001	0.001	0.001	0.001
Momentum	**b**	**-0.020*****	**-0.037*****	**-0.037*****	**-0.037*****	**-0.036*****	**-0.037*****	**-0.039*****
	se	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Bm	**b**	**0.000**	**0.001**	**0.001**	**0.001**	**0.001**	**0.000**	**0.001**
	se	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Volatility	**b**	**-0.040**	**0.140*****	**0.143*****	**0.144*****	**0.129*****	**0.137*****	**0.140*****
	se	0.025	0.020	0.021	0.022	0.026	0.019	0.020
	constant	**-0.045**	**0.088****	**0.092****	**0.084****	**0.075*****	**0.103*****	**0.097*****
	se	0.036	0.038	0.038	0.036	0.016	0.015	0.015
	N	954	954	954	954	893	941	941
	N. of groups	93	93	93	93	89	92	92
	Number of instruments	87	89	88	89	87	87	87
	Wald chi2	1.850.39	1190.71	1088.76	1038.19	1426.52	1458.62	1377.33
	z(AR1)	-2.43	-3.68	-3.66	-3.67	-3.25	-3.72	-3.72
	z(AR2)	0.32	-0.60	-0.51	-0.52	-0.32	-0.51	-0.54
	Hansen J test	70.42	73.41	71.45	72.7	70.7	70.88	70.63

Notes: Variables are defined in Table 2. Ep is earnings per share forecast for the next year over market price of a bank's common stock in June of year t. Int is net interest revenue in total operating revenue. Com is net commission and fee revenue in total operating revenue. Trad is net trading revenue in total operating revenue. Tier 1 is Tier 1 capital over total risk-weighted assets. $Zscore_{mv}$ is equal to the sum of the average return on assets (calculated by using five years of past data) and the ratio of market value of equity capital in total assets; over the standard deviation of return on assets (calculated by using five years of past data). $Zscore_{bv}$ is equal to the sum of the average return on assets (calculated by using five years of past data) and the ratio of book value of equity capital in total assets; over the standard deviation of return on assets (calculated by using five years of past data). Dependent variable is the cost of equity capital. b refers to the regression coefficient, se to robust standard error. Estimation method: 2-step GMM estimates. Instruments: lags of Average 1 to 4 as the GMM instruments and year dummies and explanatory variables as IV instruments. ***/**/* denotes statistical significance at the 1%/5%/10% level

.

Chapter 3:

Clustering and matching of US banks during the crisis

1 Introduction

Most economists agree that the financial crisis which started in the summer of 2007 (in this paper it is referred to as "the crisis") is the worst financial crisis since the Great Depression[10] (Reuters, 29.2.2009). One of its consequences was also the failure of 330 banks during 2008 and 2009 (Bexley, 2010). Another observation related to this period is the high cross-section variation of US banks' risk, measured with Z-score. This raises the intuitive question of how did banks group themselves with respect to the before and after the crisis level of risk, during the crisis change in risk and pre- and post-crisis financial statement information. Which are the variables that differ significantly between various groups of banks? Hence, the first hypothesis of this analysis is: differences can be observed between various groups of banks according to their pre-crisis financial statement information, before and after the crisis level of risk and during the crisis change in risk. Then, the second hypothesis is tested: geographic diversification had a statistically significant effect on the level and change in banks' risk during the crisis. In the last part of the analysis, the matching models approach is applied to test the third hypothesis: those banks that diversified substantially before the crisis also had a larger increase in risk during the crisis compared to those banks that did not diversify their operations much.

The reasons why various estimation techniques are employed in the analysis are the following: cluster analysis is applied because it is a helpful tool in understanding the complex nature of multivariate relationships which is being analyzed. It is worthwhile to remind the reader that the main objective of cluster analysis is to identify groups of observations (clusters) such that each cluster is as homogeneous as possible with respect to the clustering variables. Hence, first cluster technique is used in order to describe various groups of banks that formed in reality. To verify these results further regression estimation technique is applied, in order to analyze whether the relationship between banks' geographic diversification and change in (and level of) risk during the crisis is also statistically significant. Matching models analysis is applied, because even the consequences of a significant increase in the geographic diversification prior to the crisis are of interest in this analysis. Hence, this analysis is interested in estimating the causal treatment effects and matching has become a popular approach for its estimation. To put it in other words, even if it can be observed that a significant increase in banks' interstate diversification prior to the crisis and an increase in the change in banks' risk during the crisis occur together, this paper seeks to determine if a significant increase in banks' interstate diversification prior to the

[10] As the main reasons for the crisis, the bubble burst in the US housing market, subprime lending, toxic assets can be listed. It is also agreed that the US monetary policy was too expansionary in the past decade.

crisis has a causal effect, or could it be that the change in banks' risk during the crisis is caused by something else.

One of the reasons why this analysis focuses on this issue is because it provides an interesting continuation of the PhD thesis, while a more important one is that this study contributes to the existing literature in several ways. To begin with, it uses the clustering technique, which is as already argued before, a helpful tool in understanding the complex nature of multivariate relationships which is being analyzed in this study. It is also the first study that analyzes the relationship between banks' geographic diversification and their level and change in risk during the crisis. The study, which is the most related to this one is Beltratti and Stulz's (2009) one. In the study, authors analyze the stock market performance of the largest banks in the world and try to determine how it is related to bank corporate governance, its financial statement characteristics and country regulation. Other studies analyzing the crisis period focus, for instance, on corporate governance issues (Adams, 2009; Erkens et al., 2009; Kirkpatrick, 2008), change in the composition of bank capital and dividend payments (Acharya et al., 2011), bank lending (Ivashina and Scharfstein, 2008), securitization (Acharya et. al, 2009). However, none of them analyzes the relationship between banks' geographic diversification and its risk. At this point it is important to emphasis that the reason why the issue of diversification is an important one is because it is related to the optimum degree of diversification (traditional banking theory suggests a well-diversified organization, while corporate finance theory suggests that a firm should stay focused).

Next, the fact that the issue of banks' risk is an important and relevant one for several groups of stakeholders is discussed; among them are supervisors and regulators, managers of financial institutions and financial investors. There are several reasons why these groups of stakeholders find this variable as a relevant indicator of banks' performance. The regulators are interested in banks' risk because their main objective is to protect depositors and to preserve the safety and soundness of financial institutions. For managers, this issue is an important one if they hold a stake in banks' equity and more importantly because they have their human capital invested in the company they run. Financial investors are interested in banks' risk because the risk assessment has implications for investment valuation.

The empirical analysis in this paper is conducted on middle-sized and large US banks. Banks' risk is measured with Z-score while the geographic diversification is measured with the Herfindahl-Hirschman index based on deposit dispersion. The main reasons why banks' risk is chosen to be proxied with Z-score are the following. First, the paper tries to determine the effect that the crisis had on both listed and private banks, which means that variables that require market data cannot be used. Second, choosing a measure based on market data would undoubtedly result in a much smaller number of included banks and given the short time period, there would be no point in performing the regression analysis. Third, there is a number of previous research studies that use Z-score variable as a measure of risk in their empirical

analyses: Boyd et. al (2009), Camaraa et al. (2010), DeYoung and Roland (2001), Leavena and Levined (2009), Morgan and Samolyk (2003), Stiroh (2004a, 2004b, 2006).

The first conclusion of the analysis is that three groups of banks with respect to the clustering variables (the pre-crisis financial statement information, before and after the crisis level of risk and during the crisis change in risk) can be observed. In general, it holds that if before the crisis a bank had a low degree of geographic diversification, high Tier 1 ratio, low level of Z-score, low profitability and low ratio of mortgage loans over loans, then it had a smaller increase in risk during the crisis. This result can be explained by the fact that most probably these banks were engaged in profitable business lines before the crisis (which resulted in high ROAA before the crisis) and at the same time they underestimated the risk of these businesses, such as securitization (these banks had too low equity for the risk they took before the crisis). In the second part of the analysis no evidence is found to support the hypothesis that geographic diversification had a statistically significant effect on the level and change in banks' risk during the crisis. The conclusion of the last part of the analysis is that those banks that diversified substantially before the crisis had a larger increase in risk during the crisis compared to those banks which did not diversify their operations much.

In the next section, empirical literature review is discussed. Then, the empirical sample used is described whilst section 4 presents the research methodology and the empirical results. Section 5 concludes the Chapter.

2 Empirical literature review

Due to the fact that there are several papers written that focus on the crisis period and issues related to it, only those most related to this study are briefly mentioned in this section. To begin with, the study which is closely related to this one is Beltratti and Stulz (2009). The authors analyze the stock market performance of the largest banks in the world and they try to determine its relationship with bank corporate governance, its financial statement characteristics and country regulation. They find that banks which the market favored in 2006 had the worst performance during the crisis. They argue that this result supports the Tsunami explanation for the crisis. They support their argument by stating that there were certain bank characteristics that the market valued before the crisis (for instance, securitization) but at the same time the market underestimated the fact that these bank characteristics would expose banks to risks should the crisis starts. They also conclude that the bank's balance statements were more important determinants of the bank's performance during the crisis than the bank's governance and regulation.

Next, papers that find evidence of the importance which the capital had during the crisis, must be addressed. Acharya et al. (2011) investigate the composition and evolution of bank capital during the financial crisis. They conclude that the composition of bank capital shifted from

one mostly based on common equity to one based on debt (due to the fact that most of the new capital raised by banks over the period 2000-2007 was in the form of debt or hybrid claims such as preferred equity, subordinated debt) and that banks continued to pay significant dividends even during the crisis years 2007-2008. They also observe the maturity mismatch between asset and liability side of banks balance sheet. The fact is that assets and loans that banks made were mostly of the long-term type, while their non-deposit debt funding was of the short-term type. Demirguc-Kunt et al. (2010) use a sample of banks from various countries to investigate if better capitalized ones had a higher stock returns during the crisis. They find evidence that during the crisis, better capitalized banks indeed had a smaller decline in their stock market value.

There are also several papers that analyze banks' lending behavior during the crisis. By looking at loans that banks reported in their aggregate balance sheets, Chari et al. (2008) conclude that bank credit has not declined during the financial crisis, but that it has increased. Ivashina and Scharfstein (2008), who examine data on syndicated loans, show that this increase in loans during the crisis was primarily due to an increase in drawdown on pre-arranged lines of credit and not due to an increase in new loans (meaning that, banks mostly honored their prior commitments but they did not make new ones). They also show that banks that reduced their lending to a greater extent were the ones with less deposit financing and with a greater risk of credit line drawdowns. Besides, Barajas et al. (2010) analyze the slowdown of lending by large US banks over the period 2007-2009. Their main finding is that capital and not liquidity constrained lending during the crisis - more capital constrained banks had lower growth in lending during the crisis. They also find that the performance of the ten largest banks in the sample was worse than the performance of the rest of the banks included in the sample and that these large banks had smaller Tier 1 and equity over total assets ratios. Their results also show that those banks that were initially more capital constrained were the ones that increased their capital levels more.

Researchers also investigate banks' defaults during the crisis. Cole and White (2010) find - based on financial data reported by commercial banks just prior to failure - that capital, asset quality, earnings and liquidity help explain the failure of banks in 2009. To put it a bit differently, failing banks had less capital, worse asset quality, lower earnings and less liquidity. Bologna (2011) examines US bank failures over the period 2007 and 2009. His findings that those banks that had low capital ratio, profitability and asset quality had a higher probability of default in this period are consistent with previous literature. He additionally finds that a high loan to deposits ratio (meaning a higher dependence on non-deposit forms of funding) increased the banks' probability of default in this period.

Several papers deal with the importance of government support during the crisis. As Treasury (2011) reports, in September 2008, the Bush Administration proposed the Emergency Economic Stabilization Act of 2008, which was enacted into law on 3 October 2008, with the

aim of promoting the stability and liquidity of the financial system. In accordance with this law, the Troubled Asset Relief Program (TARP) was established in order to purchase and guarantee troubled assets from financial institutions. The most important part of TARP was the program that invested a total of 245 billion US dollars in banking institutions, with the goal to help banks that were under stress to begin lending again. The two most important parts of the program were capital purchase program (CPP) and targeted investment program (TIP). By providing capital to viable US financial institutions of all size classes, the main goal of the CPP was to stabilize and strengthen the financial system. In this program which was voluntary, Treasury provided 205 billion US dollars of capital to 707 financial institutions through the purchase of preferred shares. In addition to CPP investments, two banks received the TIP investments - Treasury purchased 20 billion US dollars in preferred stock from Citigroup Inc. and from the Bank of America Corporation. This was an exceptional assistance made on a case-by-case basis in order to stabilize institutions that were considered systemically important. Veronesi and Zingales (2008) estimate the costs and benefits of the intervention for the ten largest banks that were the first to receive Government help and conclude that the plan created value (they approximate it to be 71 billion – 89 billion US dollars).

Bayazitova and Shivdasani (2009) study TARP and CPP funds allocated and spent through 2009. They find that the banks which received capital injections were the ones with higher potential for systemic risk (big banks and banks that relied heavily on wholesale funding) and not the ones with distressed loan portfolios (asset quality of the selected banks was higher than the asset quality of the rejected banks). Given that capital infusion was viewed as costly by banks, healthier banks (banks with high capital ratios and strong asset quality) voluntarily turned down TARP assistance. They also examine banks' decision to exit CPP and find that large banks, with high capital ratios and better asset quality were the ones which repaid CPP infusion. They also find evidence supporting the idea that restrictions over CEO compensation played an important role in the capital infusion process. Taliaferro (2009) finds that banks used most of the received money from capital purchase program to improve their capital ratios and only a small portion of it to increase lending.

Given the importance of a sudden stop in the wholesale market before the crisis, several papers analyze this issue too. Acharya et al. (2008) and Huang and Ratnovski (2011) develop models in which they analyze the downside of bank wholesale funding (banks increasingly substitute deposits with short term wholesale funding). In their models, they explain why markets for rollover debt may experience a sudden freeze. Goldsmith-Pinkham and Yorulmazer (2010) use an event study methodology to show the limits of excessive reliance on (short-term) wholesale funding.

Furthermore, there are several papers that try to determine the main causes of the crisis. This stream of literature is not described in detail here because this paper does not investigate these

causes. However some of the papers are mentioned next for the sake of completeness. There exists a perspective which argues that the fail of corporate governance was the major cause of the crisis (see Tarraf, 2010, for literature review on the corporate governance and the financial crisis). There are also studies that apply behavioral view to explain the causes of the crisis (for example, Grosse, 2010, and Rotheli, 2010). Acharya and Richardson (2009) argue that the main reason why this crisis was much worse from the one in the year 2000 was because large institutions did not choose to spread the risk to other investors, which is the main purpose of securitization, but they kept it to themselves. Hence it can be said that, the main aim of securitization was to circumvent the capital adequacy regulation.

3 Sample

The sample includes listed and private US banks over the period 2006-2009.

In the analysis, the following databases are used:

- **Bankscope.** The Bureau Van Dijk issues the Bankscope database every month. To determine the sample for every year, the December issue for that year is used.
- **The Summary of Deposits** (SOD) is the Federal Deposit Insurance Corporation (FDIC) database. It is an annual survey of the branch office's deposits for all FDIC-insured institutions as of June 30. This survey has been conducted since 1934, but on-line data is available from 1994 afterwards.

Construction of the sample: The sample is constructed following these steps. First, US banks are identified by using the December 2006 issue of Bankscope. Next, the following banks are removed:

- with fiscal year-end total assets smaller than 5,000,000,000 US dollars,
- which are not included in the SOD database,
- with missing required accounting data in Bankscope.

The first part of the analysis, the cluster analysis only includes those banks with all necessary data available for the years 2006 and 2009 (data to calculate the Z-score must be available also for the year 2009). The same holds for the first part of the regression analysis (in which the relationship between geographic diversification and the change in banks' risk during the crisis is estimated). However, because there is less data needed for that part of the analysis, there are more banks included in the regression estimation. In the second regression estimation (in which relationship between geographic diversification and the level of banks' risk during the crisis is estimated) those banks which have data available for years 2007 and 2008 are included. The last part of the analysis requires that banks have available balance sheet information also for the year 2000.

4 Methodology

4.1 Cluster analysis

As Sharma (1996) points out, the main objective of cluster analysis is to identify groups of observations that are similar to each other with respect to the clustering variables. In other words, cluster analysis is a technique used for combining observations into groups (clusters) such that, on one hand, observations in each group are similar to each other (each group is homogeneous with respect to certain characteristics) and, on the other hand, observations of one group differ from the observations of other groups (each group should be different from other groups with respect to the same characteristics) (Sharma, 1996).

Cluster analysis can be performed by applying the following steps. Firstly, choose a measure of similarity. Secondly, select the type of clustering technique to be used - a hierarchical or a nonhierarchical technique. Thirdly, select the type of clustering method for the selected technique. Fourthly, determine the number of clusters. Lastly, interpret the cluster solution. Before describing these steps more in detail, the next section lists the variables included in the analysis.

4.1.1 Definition of variables

Due to the fact that the information on banks' geographic diversification is available for every June for all balance sheet information (such as total assets, equity), June's values are collected. In case where the second quarter balance sheet information is not available, but the bank has available data for two consecutive fiscal year ends (December) June's balance sheet value is calculated as the average, for instance, total assets for June 2006 are calculated as (December 2005 + December 2006)/2. For profit and loss items (for example, income, expenses) annual information (accounting item referring to the whole fiscal year) is collected.

Z-score

$$Z - score = \frac{\mu + k}{\sigma} \tag{1}$$

where
μ … the average return on assets, calculated by using five years of past data,
k … the equity capital over total assets, calculated as: book value of equity/total assets,
σ … the standard deviation of return on assets, calculated by using five years of past data.

The Z-score is based on accounting data and it is a measure of an individual institution's soundness. Z-score combines banks' profitability (μ), capital ratio (k) and return volatility (σ). From the formula it can be seen that it increases with banks' profitability and capital ratio and decreases with return volatility. Therefore, it can also be said that it is an indicator of the

probability of default (the situation where losses exceed equity) and that a higher Z-score implies a lower probability of insolvency.

With respect to this variable, it needs to be noted that there is no consensus among researchers about how many past values of variable to use to calculate mean and standard deviation (some researchers use five years of past data, while others use only three). Moreover some researchers use ROAE (Camaraa et al., 2010) whereas others use ROAA (Laeven and Levine, 2009). This study follows the majority of researchers, which use ROAA and calculate the ratio by using five years of past data.

In the analysis, the following two transformations of this variable are used:
$Z\text{-}score_{2009/2006}$ = (Z-score$_{2009}$ - Z-score$_{2006}$)/Z-score$_{2006}$
Ln_Zcore_t = natural logarithm of Z-score$_t$ (t= 2006-2009)

Geographic focus
To measure diversification between major geographic areas in which the bank operates, a deposit based Herfindahl-Hirschman Index (HHI_{GF}) is constructed for each bank. The index is the sum of squared deposit shares in each state that the bank operates. As HHI_{GF} rises, the bank becomes more concentrated and less diversified. In the case where the bank is fully specialized the index takes the value of 1:

$$HHI_{GF} = \sum_{i=1}^{50} \left(\frac{Deposits_{State\ i}}{Total\ \ Deposits} \right)^2 \qquad (2)$$

The information on deposit distribution across various states of the US is obtained from the Federal Deposit Insurance Corporation (FDIC), the summary of deposits (SOD) database. This database was already used before by various researchers (Morgan and Samolik, 2003; Deng and Elyasiani, 2008). SOD reports geographic segment data as annual items in its database. The data contains information about branch office deposits for all FDIC-insured institutions as of June 30. Data is available on-line from 1994 afterwards. The information is given for all federal countries in which the bank has its branches. The reporting is standardized.

Business focus
Similarly as in other papers (Acharya et al., 2006; Elsas et al., 2006; Stiroh, 2004; Stiroh and Rumble, 2006) even this one calculates banks' business focus as the Herfindahl-Hirschman index:

$$HHI_{BF} = \left(\frac{INT}{TOR}\right)^2 + \left(\frac{COM}{TOR}\right)^2 + \left(\frac{TRAD}{TOR}\right)^2 + \left(\frac{OTI}{TOR}\right)^2 \quad (3)$$

where INT … net interest revenue,
COM … net commission and fee revenue,
TRAD … net trading revenue,
OTI … all other revenue,
TOR … total operating revenue, equal to the sum of the absolute values of INT, COM, TRAD and OTI.

As Elsas et al. (2006) suggest, absolute values are used in the calculation. The Herfindahl-Hirschman index increases with bank concentration. In the case where the bank is fully specialized and generates all revenues from one revenue stream, the index takes the value of 1, while in the case where the bank is fully diversified across all four revenues streams (business areas), the index takes the value of 0.25.

Size = natural logarithm of book value of total assets (in US dollars)
Cost-to-income-ratio = operating expenses/operating income
Equity = book value of equity/total assets
Liquidity ratio = liquid assets/total assets
Mortgage loans = (residential mortgage loans + other mortgage loans)/loans
Corporate loans = corporate & commercial loans/loans
Deposits = deposits/total assets

Moreover, the following variables too are being reported even if they are not included in the clustering estimation since they can help interpret banks' clusters even better. These variables cannot be included in the cluster analysis because they are highly correlated with some other variables already included in the cluster analysis.
Tier 1 = Tier 1 capital/total risk-weighted assets
Loans = loans/total assets
Net interest margin = (interest income – interest expenses)/average of preceding and current year earnings assets
ROAA = net income/average of preceding and current year total assets
ROAE = net income/average of preceding and current year total equity
MM and ST funding = money market and short term funding/total assets
Loans over deposits = loans/deposits
LLprov_loans = (loan loss provisions$_{2007}$+loan loss provisions$_{2008}$)/((loans$_{2007}$ + loans$_{2008}$)/2)
Banks' credit risk is being captured with the amount of reserves that managers set up during the crisis to cover unexpected future losses on loan defaults. Because the values of credit risk variable in the two crisis years are highly correlated, including them both in the analysis is not

recommendable. In order to overcome this problem, the two values are summed up to get the number that refers to the whole crisis period.
Interest Revenue = net interest revenue/total operating revenue
Commission and fee revenue = net commission and fee revenue/total operating revenue
Trading revenue = net trading revenue/total operating revenue
Other revenue = all other revenue/total operating revenue

4.1.2 Step one in cluster analysis: *select a measure of similarity*

All clustering algorithms require some type of measure to assess the similarity of a pair of observations or clusters (Sharma, 1996). The most frequently used similarity measure is a distance measure, and there are a number of them. Some of the most frequently cited are the following:

The *Minkowski distance* between points *i* and *j* in *p* dimensions is given by:

$$D_{ij} = \left[\sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^n \right]^{1/n} \quad (4)$$

$n = 1, 2, \ldots$
D_{ij} … distance between observations *i* and *j*,
p … number of variables,

A value of n = 2 in the equation gives Euclidean distance and a value of n = 1 results in Manhattan distance.

The *Euclidean distance* between points *i* and *j* in *p* dimensions is given by:

$$D_{ij} = \left[\sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^2 \right]^{1/2} \quad (5)$$

D_{ij} … distance between observations *i* and *j*,
p … number of variables.

The *squared Euclidean distance* between points *i* and *j* in *p* dimensions is given by:

$$D_{ij}^{\ 2} = \sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^2 \quad (6)$$

$D_{ij}^{\ 2}$ … squared distance between observations *i* and *j*,
p … number of variables.

The *Manhattan distance* between points *i* and *j* in *p* dimensions is given by:

$$D_{ij} = \sum_{k=1}^{p} \left| X_{ik} - X_{jk} \right| \quad (7)$$

D_{ij} … distance between observations *i* and *j*,

p … number of variables.

The final choice among them depends on the data and the type of variables collected. The following facts are reasons why the Euclidean distance is an appropriate choice in this analysis. First of all, the analysis applies standardization methodology, which is more robust. Second of all, the variables used are relatively weakly correlated once standardized. Furthermore, such variables are tried to being collected to avoid multicollinearity.

One of the problems of using a distance measure as a similarity measure is that it is not scale invariant, which means that the distance between observations could change with a change in scale (Sharma, 1996). To overcome this problem, ratios are being used instead of absolute values of financial statement information (the only non-ratio variable included in the analysis is the size variable, which is measured with the natural logarithm of total assets). The data is also standardized. The main reason for standardization is that those variables with a larger scale would have had a greater impact in each cluster than other variables and hence would have dominated and potentially biased the results. To transform variables z scores transformation is chosen. In this case values are standardized with a mean of 0 and a standard deviation of 1. A z score quantifies the original score in terms of the number of standard deviations that that score is from the mean of the distribution.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (8)$$

x_i … an observation I,

μ … mean,

σ … standard deviation.

4.1.3 Step two in cluster analysis: *select clustering technique*

Clustering algorithm techniques can be divided into two main groups: nonhierarchical and hierarchical techniques. The hierarchical technique is applied in this analysis, because the final number of clusters is not known in advance. This is also the main advantage of the hierarchical clustering methods - they do not require a priori knowledge of the number of clusters or the starting partition before the technique can proceed to cluster observation. Among the disadvantages of hierarchical methods, the main one is that once an observation is assigned to a cluster it cannot be reassigned to another one. It can be added that one of the disadvantages of non-hierarchical clustering algorithms is that they are very sensitive to the

initial partition. As a number of starting partitions can be used, Sharma (1996) notices that the final solution could result in local optimization of the objective function.

The output from hierarchical techniques can be represented by a dendrogram. A dendrogram is used as a support in deciding the number of clusters. It illustrates the mergers or divisions, which have been made at successive levels and the distance between clusters.

Hierarchical clustering methods start with the individual objects (banks in this analysis) and initially there are as many clusters as objects. In this analysis, hierarchical clustering methods start by placing each bank in its own cluster. The most similar objects are first grouped - the two closest banks are fused into a cluster according to their similarities (selected linkage method). Next, either a new bank is added to the cluster, or another two-bank cluster is formed. This process continues until all subgroups of banks are fused into a single cluster.

4.1.4 Step three in cluster analysis: *select the clustering method*

Next, the rule to be used for determining the distance or similarity between clusters consisting of more than one subject must be determined. A number of different rules or methods have been suggested for computing distances between two clusters. In fact, the various hierarchical clustering algorithms or methods differ primarily with respect to how the distances between the two clusters are computed. Some of the popular methods are: centroid method, nearest-neighbor or single-linkage method, farthest-neighbor or complete-linkage method, average-linkage method and Ward's method.

Sharma (1996) makes the following points about various techniques. First of all, hierarchical methods are susceptible to a chaining effect. This means that observations are sometimes assigned to existing clusters rather than being grouped in new clusters. In general, this problem is more susceptible in the nearest neighbor. Second of all, compared to the single-linkage method, the complete-linkage method (farthest-neighbor method) is less affected by the presence of noise or outliers in the data. Third of all, the Ward's method tends to find clusters that are compact and nearly of equal size and shape. Based on the last argument, Ward's method is used in this analysis. It is necessary to add that this method, rather than computing the distance between clusters, forms clusters by maximizing within-clusters homogeneity. Therefore, it can be argued that the within-cluster sum of squares is used as the measure of homogeneity, as this method tries to minimize the total within-cluster sum of squares. (Sharma, 1996)

4.1.5 Step four in cluster analysis: *determine the number of clusters*

To determine the number of clusters to explore more in detail, it is helpful to look at the dendrogram or tree diagram. The statistical package SPSS is used to produce it. A dendrogram graphically represents the sequence of clustering by displaying the observations, the sequence of clusters and the distances between the clusters (Härdle and Simar, 2007). It

plots distance between clusters on the horizontal axis and sample units on the vertical axis. Large distances indicate the clustering of heterogeneous groups.

The following type of information can be obtained from dendrogram: weight, compactness and distinctness of each cluster. Weight refers to the importance of each cluster and it is represented by the share of observations that fall within each cluster (a number of leaves that that branch of the dendrogram leads to) (ArcObjects, 2002). Compactness refers to similarity between the elements of a cluster and it represents the minimum distance at which the cluster comes into existence (ArcObjects, 2002). Distinctness refers to how different a cluster is from its closest neighbor and it is measured by the distance along the horizontal axis, from the point at which the clusters come into existence to the point at which they aggregate into a larger cluster. (ArcObjects, 2002).

The dendrogram is displayed in Figure 1. As the aim of classification is to choose clusters that are as compact and as distinct as possible, it can be concluded that in this example, there are three distinct clusters that should be explored more in detail.

4.1.6 Step five in cluster analysis: *interpret the cluster solution*

In order to examine how clusters evolve over time a clusters' analysis is performed by using balance sheet information prior (for the year 2006) and after the crisis (for the year 2009)[11]. The results are reported in Tables 2a and 2b. The last columns of these Tables report analysis of variance (ANOVA), which tests the hypothesis that several means are equal. Based on this test, it can be concluded that there are significant differences between groups with respect to most of the variables used in cluster analysis. Additionally Tables 2c and 2d report results of the t-test (for clusters obtained by using the balance sheet information for the year 2006 and 2009 respectively), which provides the information about the variables that characterize differences between various clusters. Hence, which are the variables that differ significantly between clusters 1 and 2, clusters 1 and 3, and clusters 2 and 3[12].

The following variables have statistically significant mean difference between clusters 1 and 2 for both 2006 and 2009: geographic and business focus, net interest revenues, net commission and fee revenues, size, mortgage and corporate loans over loans, loan loss provisions over

[11] Moreover, a principal component analysis is performed. This one shows that there is a correlation between variables but not one that is big enough that variables used in the analysis could be substituted with a couple of components obtained with the principal component analysis.

[12] Furthermore, the author also checks if clusters obtained by using balance sheet information for the years 2006 and 2009 give similar results for both years. Hence, she tests if the balance sheet information for the year 2006 (2009) for clusters obtained by using the balance sheet information for the year 2006 differs significantly from balance sheet information for the year 2006 (2009) for clusters obtained by using the balance sheet information for the year 2009. The results, which are not reported, show only 3 statistically significant differences; 2 in clusters 2 and 1 in cluster 3.

loans and Zscore after the crisis. Therefore, it can be said that banks from cluster 1 are much more diversified over business activities and geographic regions than banks from cluster 2; they earn less net interest revenues and more net commission and fee revenues, they are also bigger, have a higher share of corporate loans in loans and smaller share of mortgage loans in loans, they set up a larger amount of loan loss provisions during the crisis, and they also had a lower Zscore after the crisis. Another interesting observation is that the two clusters differ with respect to the variable loans over total assets for the year 2009. This means that before the crisis, both groups of banks had on average a similar value of this ratio, but after the crisis banks from cluster 1 decreased the amount of loans in their asset structure whilst those from cluster 2 increased it.

Variables that have statistically significant mean difference between clusters 1 and 3 for both 2006 and 2009 are the following: geographic and business focus, net interest revenues, net commission and fee revenues, size, cost to income ratio, ROAA and ROAE, Tier 1 ratio, corporate loans over loans, Zscore before the crisis and change in Zscore during the crisis. This suggests that banks from cluster 1 are much more diversified over business activities and geographic regions than banks from cluster 3; they also earn less net interest revenues and more net commission and fee revenues, they are bigger, have a higher cost to income ratio, had a higher profitability before the crisis and lower profitability after the crisis, have a smaller Tier 1 ratio and have a bigger share of corporate loans in loans. They also had a higher Zscore before the crisis and a larger increase in risk during the crisis.

The following variables have statistically significant mean difference between clusters 2 and 3 for both 2006 and 2009: mortgage loans over loans, Zscore before the crisis and change in Zscore during the crisis, which implies that banks from cluster 2 have more mortgage loans over loans than banks from cluster 3. They also had a higher Zscore before the crisis and a larger increase in risk during the crisis.

Based on this discussion, it can be concluded that clusters 1 and 3 differ with respect to a number of variables and that even clusters 1 and 2 differ with respect to several variables, while clusters 2 and 3 differ with respect to only a few of them. Based on this, more observations are expected to be seen switching from cluster 2 in 2006 to cluster 3 in 2009 than from cluster 1 in 2006 to clusters 2 or 3 in 2009. Looking at Table 2e it can be concluded that this observation is correct, as 8 observations went from cluster 2 in 2006 to cluster 3 in 2009, which is the largest number of changes between clusters. There were, however, also 5 observations that went from cluster 3 in 2006 to cluster 1 in 2009. But overall, it can be noted that clusters are fairly stable over time and that most of the changes are due to the changes over time and not due to changes in units within clusters.

Next, the analysis tries to give an economic interpretation of the three clusters obtained. For the rest of the section it refers to the clusters as defined in cluster analysis, which uses data for balance sheet information for the year 2006 and results reported in Table 2a.

The first group is made of 40 large banks that are the most geographic and business diversified. These banks had the lowest relative net interest income (0.590 in 2006) and the highest relative commission and fee income (0.205 in 2006) and relative trading income (0.022 in 2006) of all three groups of banks before the crisis. They also had a high cost to income ratio (0.623 in 2006) and the highest ROAA (0.013 in 2006) and ROAE (0.138 in 2006) before the crisis. This implies that before the crisis these banks ran their operations profitably but not very efficiently. Moreover, this group of banks had a high ratio of mortgage loans over loans (0.190 in 2006), the highest ratio of corporate loans over loans (0.465 in 2006), the highest loans over deposits ratio (0.830 in 2006) and the lowest Tier 1 ratio (0.094) before the crisis. This is the group of banks that had the highest negative change in Z-score during the crisis, -0.739 (average natural logarithm of Z-score was 2.662 in 2006 and 0.743 in 2006). The reasons for this negative change in Z-score are that first of all, before the crisis banks forming this group had a low equity over total assets ratio (0.100 in 2006) and secondly, they also suffered a substantial drop in average ROAA during the crisis (5-year average ROAA dropped from 0.014 in year 2006 to 0.006 in year 2009). The intuition to explain this substantial increase in risk of these banks during the crisis can be found in their securitization activities and their exposure to the business sector.

Based on the observation that these banks had the highest relative noninterest income before the crisis, it can be assumed that this group was the most involved in securitization activities of all three groups of banks. By removing a high concentration of risk from the balance sheets of financial institutions and placing small concentrations of risk to a large number of investors, the purpose of securitization should be to spread risk. However, Acharya and Richardson (2009) argue that over the period 2003-2007 the main goal of securitization was not this one, but to circumvent the capital-adequacy regulation. According to them, the risk remained concentrated in the financial institutions and this practice allowed banks to over leverage, which magnified the risk they took. The point is that through securitization banks still make loans but as they sell them off, these loans disappear from their balance sheets and banks can avoid setting up additional capital reserves. However, banks, which set up the conduit, had to provide guarantees that they will provide backup credit lines if the conduit will not be able to pay back investors. This means that these banks did not remove the risk even if it was not shown in their balance sheet. Acharya and Richardson (2009) report that when the crisis hit, only 4.3 percent of the 1,250 billion US dollars in asset-backed securitized vehicles loss was structured to remain with investors. Additional evidence for this explanation is a very low Tier 1 ratio of this group of banks before the crisis, meaning these banks were indeed trying to have as low regulatory capital as possible. Acharya and Richardson (2009) add that another common bank practice before the crisis was to invest in AAA-rated tranches

of the securitized products. The reason for this decision was low capital requirement for these assets, because of their high ratings. Both of these bank activities made the crisis much worse than it would have been if the banks transferred risk to other investors rather than keeping it for themselves.

Another reason why this group of banks had a substantial increase in its risk was also its exposure to the business sector which was facing the worst downturn since the Great Depression. It can be added that NBER determined that the crisis began in the US in December 2007 and that a trough in business activity occurred in June 2009, which means that the recession lasted 18 months, and that this is the longest recession since World War II. Obviously there was a decline in real GDP growth and the percentage change from the preceding period became even negative in 2009 (source: Bureau of Economic Analysis). An increase in the unemployment from 4.6% in 2006 and 2007 to 9.6% in 2010, which is the highest rate since 1983 (source: Bureau of Labor Statistics), also produced a reduction in domestic demand implying a decline in revenues for most of the firms.

In the second group of banks 33 moderately geographically and business diversified banks can be observed. These are small banks, which had a high relative net interest income (0.746 in 2006) and the lowest relative commission and fee income (0.085 in 2006) of all the three groups of banks before the crisis. They also had a Tier 1 ratio of 0.123, the highest deposits over total assets ratio (0.716 in 2006) and the highest ratio of mortgage loans to loans (0.728 in 2006) before the crisis. This group of banks had a large negative change in Z-score during the crisis, -0.546 (average natural logarithm of Z-score was 2.976 in 2006 and 1.550 in 2009). The reason for this negative change in Z-score being a high drop in average ROAA during the crisis (5-year average ROAA dropped from 0.012 in year 2006 to 0.008 in year 2009).

One of the main reasons why this group of banks had a substantial increase in risk during the crisis was its significant credit exposure to the housing sector. It is important to remember that the trigger of the financial crisis was the burst of the US housing bubble, which peaked in 2005-2006. Over the period June 1997-June 2006 the price of an average American house increased by 128% according to the S&P/Case-Schiller national home price index (see Figure 2). Given this price appreciation a lot of homeowners took out a second mortgage to finance their consumer spending. Therefore, it can be said that in general consumers took advantage of low interest rates (FED lowered the federal funds rate target from 6.5% to 1.0% from 16 May 2000 to 25 June 2003, source: FED) to borrow large amounts of money. Furthermore, given the optimism, a lot of mortgage lenders gave loans to borrowers without carefully examining if they will be able to pay them back (Bush, 24.9.2008). Lindsey (2007) reports that by 2006 half of first time home buyers had a down payment of only 2%, while just 15 years ago this number was 20%. And it adds that 40% of all first time home buyers put no down payment or took a mortgage that was bigger than the cost of the house. As interest rates begin to rise (FED increased the federal funds rate target from 1.0% to 5.25% from 25 June

2003 to 29 June 2006; source: FED), and housing prices started to drop (as can be seen from the Figure 2 the average American housing prices declined by 22% from June 2006 to September 2008), refinancing became more and more difficult and borrowers began to default. By September 2009, 14.41% of all American mortgages outstanding were either delinquent or in foreclosure (MBA, 19.11.2009).

The third group consists of 25 middle-sized banks that are geographically focused. Before the crisis this group of banks had a low cost-to-income ratio (0.564 in 2006), low net interest margin (0.032 in 2006), the lowest ROAA (0.010 in 2006) and ROAE (0.105 in 2006) and the highest Tier 1 (0.166 in 2006) of all groups. They also had the highest ratio of money market and short term funds over total assets (0.114 in 2006) before the crisis. This group of banks suffered the smallest change in Z-score during the crisis, 0.030 (average natural logarithm of Z-score was 2.020 in 2006 and 1.233 in 2009). One of the reasons for this is that banks from this group had high equity over total assets ratio (0.111 in 2006) before the crisis. In addition, it can be said that because this was a liquidity crisis these banks were in a better position than the other two groups of banks as they had more highly liquid assets in their balance sheet.

From Table 2a, it can also be noticed that all bank groups experienced a significant reduction in profitability during the crisis. In addition, after the crisis, banks from group 1 were the ones which decreased loans over total assets, while the other two groups of banks increased this ratio. Banks from group 1 also increased their Tier 1 and equity over total assets.

Based on the discussion above and by looking at the Table 2c, it can be said that the variables that differ the most between the two groups of banks (groups of banks 1 and 2), which suffered the highest increase in risk during the crisis and the group of banks (group of banks 3) that had the smallest increase in risk during the crisis are: geographic diversification, profitability, Tier 1, mortgage loans and level of Z-score before the crisis. In other words, if before the crisis, a bank had lower geographic diversification, lower profitability, higher Tier 1 ratio, lower ratio of mortgage loans over loans and lower level of Z-score, it also had a smaller increase in risk during the crisis.

As it was already discussed in the literature review, the US government decided to help financial institutions under stress so that they would start lending again. This was achieved by injecting a total of 245 billion US dollars in banking institutions (205 billion US dollars in capital purchase program and 40 billion US dollars in targeted investment program in which the government helped Citigroup Inc. and Bank of America Corporation). This issue is relevant also for this analysis. Hence, first, the number of banks from each cluster that received capital injection is determined. As can be seen in Table 2f, 29 out of 40 banks in cluster 1 received it, while there were less than half of the banks from cluster 2 that received funds and only 4 out of 25 banks from cluster 3 that received help. Therefore, it can be said that this result is consistent with the observation about the average change in banks' risk

during the crisis for each cluster - most of the banks from the cluster that had the highest increase in risk during the crisis (cluster 1) received capital injection. There was also a number of banks from cluster 2 (which also suffered a notable increase in banks' risk during the crisis) that received help, while there were only a few banks from cluster 3 that received help (this group of banks was the least affected by the crisis).

Another observation that can be made is that a lot of banks from cluster 1 (19 out of 29) and a majority of banks from clusters 3 (3 out of 4) repaid received funds by the end of June 2011. It can be expected that large banks did so because of restrictions on executives compensations (Bayazitova and Shivdasani, 2009, find that restrictions of executive compensation had an important role in banks' capital purchase program exiting decisions), while banks from cluster 3 did so as they were healthy enough to find cheaper sources of financing once financial markets stabilized (all three banks repaid the capital received already in 2009). Furthermore, comparison of balance sheet characteristics for the year 2009 of banks that repaid all the capital to the Treasury by the end of June 2011 and those who did not is done. These results are not reported in detail, nevertheless it can be said that banks that repaid all funds received were different from banks that repaid only part or no funds at all with respect to the following characteristics. Banks that repaid all funds received were bigger, more profitable (had a higher ROAA and ROAE ratios), more diversified, had a higher ratio of liquid assets over total assets, had a higher level of Zscore and a smaller change in Zscore during the crisis. Contrary to the findings in Bayazitova and Shivdasani (2009), a statistically significant difference in the capital ratios between the two groups of banks is not found in this paper. There is also no statistically significant difference between the three clusters in the ratio of funds received over the total assets in the year they received funds. This ratio is on average 2.25% for all banks in the sample that received capital purchase program funds (results do not change much if also targeted investment program funds are included in calculation).

Next, the similarity of these results to the results of related studies is discussed. First these results are similar to the ones in Beltratti and Stulz (2009). Furthermore, results in this paper are also consistent with those papers which find that those banks that did not increase their lending during the crisis were the ones that had the highest leverage before the crisis and had to increase their capital the most during the crisis (Acharya et al., 2009). In addition, results in this paper are also consistent with papers that find that performance of the largest banks was the worst (Barajas et al., 2010), that large banks had the highest leverage and that they had to increase their capital ratio significantly during the crisis (Barajas et al., 2010; Demirguc-Kunt et al., 2010). Moreover as Ivashina and Scharfstein (2008), even this paper finds that banks that had the lowest ratio of deposits over total assets, were the ones that reduced their ratio of loans over total assets during the crisis, while those banks that had the highest ratio of deposits over total assets were the ones that increased their ratio of loans over total assets the most during the crisis.

Findings of this paper also indicate that those banks that suffered the most during the crisis were the ones that had high loans to deposit ratio, hence, those banks that relied heavily on non-deposits sources of funding (as in Bologna, 2011); were inefficient, big and had high loan over deposit ratio (as in Igan and Pinheiro, 2009). Igan and Pinheiro's (2009) explanation about the specialization of mortgage banks (because these banks specialize in real estate lending, they gain the ability to make better loans in this market) can also explain smaller loan loss provision during the crisis that banks from group 2 made comparing to loan loss provisions made by the other two groups.

Findings in this paper are also consistent with findings about government help. The majority of government funds went to banks from cluster 1, which includes the largest and the most systemically important banks (similar finding also in Bayazitova and Shivdasani, 2009). The percentage of total funds received by this class of banks is 73% if only capital purchase program funds are considered and 77% if also targeted investment program funds are considered. It can also be noted, from Table 2a, that banks from cluster 1 did not increase their lending during the crisis but they did improve their capital ratios (which is consistent with findings in Taliaferro, 2009).

4.2 Regression analysis

4.2.1 Geographic focus (diversification) and change in banks' risk during the crisis

In the first part of regression analysis, the relationship between geographic focus and a change in banks' risk is examined by using OLS estimator. This estimation method is used because the sample period consists of only one time period. More specifically, the following regression model is estimated:

$$Zscore_{it} = \alpha_0 + \alpha_1 Geog\_focus_{it} + \sum_{n=2}^{N} \alpha_n X_{nit} + \varepsilon_{it} \qquad (9)$$

$Zscore_{it}$ … change in bank risk over the period 2006 to 2009, as described in section 4.1.1

$Geog\_focus_{it}$ … the Herfindahl-Hirschman index of geographic focus, as described in section 4.1.1

X_{nit} … control variables

ε_{it} … error term

The following *control variables* are included in the model:

- *Business focus:* the reason for including business focus in the regression is because Bodnar et al. (1998) show that by not including both types of diversification in the model together, there emerges an omitted variable problem.
- *Size:* defined as the natural logarithm of total assets, banks' size needs to be controlled for as the analysis includes banks of different size groups (middle-sized and large banks).

Given the finding in the first part of the analysis (during the crisis, large banks had a significant increase in risk) a negative relationship between this variable and an increase in Z-score during the crisis is expected to be found.

- *Leverage*: defined as equity over total assets. The reason for including this variable is because Berger and Bouwman (2009) show that capital helps banks to survive banking crises.
- *Credit risk:* defined as the ratio between the sum of loan loss provisions that a bank made in 2007 and 2008 and the average of loans that a bank had in 2007 and 2008. Loan loss provisions are the amount that management reserves to cover unexpected future losses on loan defaults; thus a bank making a small number of risky loans should have less loan loss provisions compared to a bank taking higher risks. Hence, a negative relationship between this control variable and an increase in Z-score during the crisis is expected to be found.
- *Level of risk before the crisis*: defined as a natural logarithm of Z-score.
- *Cost-to-income ratio:* defined as operating expenses over operating income. This ratio measures banks' operating efficiency. The lower the ratio, the more efficient a bank is. The ratio is used as there might be a positive relationship between bank cost efficiency and banks' risk, meaning banks which are less efficient are also more risky. There are several empirical studies supporting this intuition. Berger and Humphrey (1992) and Wheelock and Wilson (1995) find that even efficiency can explain banks' failure. Moreover, Barr and Siems (1994) find that failing banks have less efficient management. Furthermore, Berger and De Young (1997) find that cost efficiency is negatively correlated with nonperforming loans; one of the reasons why this might be so is because of poor loan portfolio management. Inefficient managers may be less able to choose loans with positive net present value, to correctly evaluate the value of collateral pledged against the loans, or to notice when loans are going bad and to take remedial action. Similarly, Kwan and Eisenbeis (1997) find a positive effect of inefficiency on the bank's risk. They find empirical support for the argument that inefficient banks are more likely to engage in riskier lending. Their reasoning follows the intuition that, because high costs reduce expected rates of return on equity, ceteris paribus, this may induce high-cost banks to increase expected return by undertaking more risky activities, which at the same time increases their probability of failure. Therefore, a negative relationship between this control variable and an increase in Z-score during the crisis can be expected.
- *Deposits over total assets:* this variable is included in the model to capture the composition of liabilities. Deposits are considered a stable source of financing; so if this is true, those banks which relied more heavily on deposit financing before the crisis were better prepared for the crisis. Hence, a positive relationship between this control variable and an increase in Z-score during the crisis is expected to be found.
- *Liquidity ratio*: this variable is defined as liquid assets over total assets. Due to the fact that this was also a crisis of liquidity, it is necessary to include also this control variable, because it can be expected that those banks which had more liquid assets before the crisis

were better prepared for it. Therefore, a positive relationship between this control variable and an increase in Z-score during the crisis is expected to be found.

With the aim of making geographic (HHI_{GF}, equation 2) and business focus (HHI_{BF}, equation 3) variables comparable and estimation results easier to interpret, these two variables are standardized. Standardized ratios are not used in cluster analysis too because the data is standardized before making the cluster analysis.

$$Geog\_focus = \frac{HHI_{GF} - 1/50}{1 - 1/50} \tag{10}$$

$$Bus\_focus = \frac{HHI_{BF} - 1/4}{1 - 1/4} \tag{11}$$

Both ratios take the values from 0 to 1.

All variables, except change in Z-score and loan loss provisions, are taken for the year 2006.

4.2.2 Geographic focus (diversification) and level of banks' risk during the crisis

The relationship between geographic focus and a level of banks' risk is investigated as an additional robustness check. An OLS estimator is applied by including a year dummy – the analysis consists of two year time period: 2007-2008. The following regression model is estimated:

$$\ln\_Zscore_{it} = \alpha_0 + \alpha_1 Geog\_focus_{it} + \sum_{n=2}^{N} \alpha_n X_{nit} + \varepsilon_{it} \qquad t = 2007, 2008 \tag{12}$$

$\ln\_Zscore_{it}$ … level of bank risk (natural logarithm of Z-score), as described in section 4.1.1

$Geog\_focus_{it}$ … the Herfindahl-Hirschman index, reflecting the degree of geographic focus (diversification), as described in section 4.1.1

X_{nit} … control variables

ε_{it} … an error term

The following *control variables* are included in the model:

- *business focus*,
- *size* (natural logarithm of total assets),
- *leverage* (equity over total assets),
- *credit risk* (loan loss provisions over loans),

- *cost-to-income ratio* (operating expenses over operating income),
- *deposits over total assets*,
- *liquidity ratio* (liquid assets over total assets),
- *time dummy.*

The intuition for their use is similar as in the previous model.

4.2.3 Results

Looking at Table 3, it can be seen that the average change in Z-score during the crisis (over the period 2006-2009) was -0.547, and that even the standard deviation, which was 0.589, was quite high. This means that there exists a significant variation in change of bank risk during the crisis for US banks. The same also holds for the level of bank risk during the crisis. The average of the natural logarithm of Z-score over the period 2007-2008 was 1.718, while the standard deviation was 1.139 (see Table 6). Since descriptive statistics for independent variables are discussed already in section 4.1.6, they are not discussed here again.

The main finding of the first regression is that those banks that before the crisis had high equity over total assets and high liquid assets over total assets, had a positive change in Z-score (an increase in bank safeness) during the crisis. Furthermore, banks that before the crisis had high cost-to-income ratio (inefficient banks) and Z-score, that were large and that made a large amount of loan loss provisions during the crisis, had a negative change in Z-score (an increase in banks' risk) during the crisis.

The main finding of the second regression is that those banks that during the crisis had high equity over total assets and high deposits over total assets had an increase in Z-score (an increase in the level of banks' safeness) during the crisis. Moreover, banks that made a lot of loan loss provisions during the crisis, had high cost-to-income ratio (inefficient banks) and had focused business operations, experienced a decrease in Z-score (an increase in level of banks' risk) during the crisis.

Another conclusion that can be drawn, based on the regression results, is that regression coefficient of geographic focus variable is not statistically significant neither in the first nor in the second regression.

Results do not change substantially if the equity over total assets ratio is replaced with Tier 1 ratio. However, in this case, r^2 decreases, and in the second regression model this variable is not statistically significant anymore. This is the reason why the equity over total assets ratio is preferred to be used in this and also the cluster analysis.

4.3 Matching models

4.3.1 Introduction

Matching has become a popular approach in estimating causal treatment effects. This approach can be applied to situations where researchers can identify a treatment, a group of treated individuals and a group of untreated individuals. This paper analyzes the effect of an increase in banks' interstate diversification before the crisis on the change in banks' risk during the crisis.

One of the main problems of the matching approach is that of missing data. This arises because researchers would like to know the difference between the participants' outcome with and without treatment. However, at any time, participants can be only in one of the two potential states but not in both of them. To put a bit differently, both outcomes for the same participants cannot be observed at the same time. This problem is known as selection bias, and one of its possible solutions is the matching approach. The basic idea of the matching approach is to find in a group of untreated individuals those who are similar to the participants in all relevant pre-treatment characteristics. By doing so, differences in outcomes of the selected and the control group can be attributed to the treatment.

Rosenbaum and Rubin (1983) suggest using balancing scores. The propensity score is one of the possible balancing score and is also the one applied in this analysis. The propensity score can be defined as a probability of participating in a treatment given observed characteristics, while matching procedures, which are based on balancing score, are defined as propensity score matching.

Furthermore, it is important to point out that this methodology was recently applied also to banking literature. Sarkisyan et al. (2009) apply propensity score matching approach in their analysis of whether banks improve their performance through the use of securitization.

4.3.2 Basics intuition of the propensity score matching approach

If a researcher wants to estimate the impact of a treatment on the outcome of an individual, he needs to speculate about how this individual would have performed had he not received the treatment. In this example it can be said that if a researcher wants to estimate the causal effect of geographic diversification, he needs to determine what the performance of diversifying banks would have been, if they had not diversified. Hence, let the treatment indicator be denoted as D_i where i is equal to 1 if an individual receives treatment and 0 if he does not. Similarly, let the outcomes be denoted as Y_1 and Y_0, where 1 is associated with received treatment and 0 with not received treatment.

The outcome observed for an individual can be defined as:

$$Y_i = DY_i(1) + (1-D)Y_i(0) \quad (13)$$

And the gain from participating in a treatment (the treatment effect for an individual i) can be defined as:

$$\tau_i = Y_i(1) - Y_i(0) \quad (14)$$

If both outcomes for the same individual could be simultaneously observed, there would be no evaluation problem because the gain from participating could be constructed for everyone. But due to the fact that this is not possible, because only one of the potential outcomes is observed for each individual, estimating the individual treatment effect is not possible and one has to concentrate on (population) average treatment effects. As is usual in research, this analysis calls the unobserved outcome, counter-factual outcome.

The most commonly used evaluation parameter is the average treatment effect on the treated (ATT), which can be defined as:

$$\tau_{ATT} = E(\tau|D=1) = E[Y(1)|D=1] - E[Y(0)|D=1] \quad (15)$$

Given that the counterfactual mean for those being treated ($E[Y(0)|D=1]$) cannot be observed, a proper substitute for it must be found in order to estimate ATT. Using the mean outcome of untreated individuals ($E[Y(0)|D=0]$) in non-experimental studies can be problematic, because it can be that components which determine the treatment decision also determine the outcome variable of interest. This means that the outcomes of individuals from the treatment and the comparison group would differ even in the absence of the treatment leading to a self-selection bias. For ATT this one can be noted as the difference between the left hand side of equation (15) and τ_{ATT}. (Caliendo and Kopeinig, 2008)

$$E[Y(1)|D=1] - E[Y(0)|D=0] = \tau_{ATT} + E[Y(0)|D=1] - E[Y(0)|D=0] \quad (16)$$

The true parameter τ_{ATT} is only identified, if the following holds:

$$E[Y(0)|D=1] - E[Y(0)|D=0] = 0 \quad (17)$$

While in social experiments, where assignment to treatment is random, this is ensured, in non-experimental studies one has to invoke some identifying assumptions to solve this problem. The assumptions are the following.

A1. Conditional Independence Assumption:

$$Y(0), Y(1) \perp\!\!\!\perp D \mid X, \quad \forall X \qquad \text{(A1)}$$

where $\perp\!\!\!\perp$ denotes independence

This means that conditional on X outcomes of non-participant have the same distribution as outcomes of participants would have if they had not participated in the treatment. Therefore, the following holds:

$$E[Y(0) \mid X, D=1] = E[Y(0) \mid X, D=0] = E[Y(0) \mid X] \qquad (18)$$

And the missing contrafactual mean can be constructed from the outcome of untreated. (Heckman et. al, 1997).

This assumption can also be explained by saying that there are no differences between the treated and untreated individuals after conditioning on X, which means that any systematic differences in outcome can be attributed to the treatment effect.

Rosenbaum and Rubin (1983) suggest using balancing scores. The authors show that if potential outcomes are independent of the treatment conditional on covariates X, they are also independent of treatment conditional on a balancing score b(X). One possible balancing score is the propensity score $P(D=1 \mid X) = P(X)$, which is the probability for an individual to participate in a treatment given his observed covariates X. The conditional independence assumption based on the propensity score can be written as:

$$Y(0), Y(1) \perp\!\!\!\perp D \mid P(X), \qquad \forall X$$

A2. Common Support:

A further requirement needed is the common support or overlap condition:

$$0 < P(D=1 \mid X) < 1, \text{ for all X} \qquad \text{(A2)}$$

This one ensures that individuals with the same X values have a positive probability of being either participants or non-participants.

The assumptions (A1) and (A2) together are called "strong ignorability" (Rosenbaum and Rubin, 1983), and if they hold, the mean outcome observed for the matched non-participating group can be substituted for the missing contrafactual mean for the participant (Smith and Todd, 2005).

Estimation Strategy: Given the two assumptions, the propensity score matching estimator for ATT can be written as:

$$\tau_{ATT}^{PSM} = E_{P(X)|D=1}\{E[Y(1)|D=1, P(X)] - E[Y(0)|D=0, P(X)]\} \quad (19)$$

Therefore, it can be said that the propensity score matching estimator is a mean difference in outcomes over the common support, weighted by the propensity score distribution of participants (Caliendo and Kopeinig, 2008) or in our case, the average diversification effect is estimated as the difference between the mean performance gain of the banks that had a large growth in geographic diversification before the crisis, after their increase in diversification, and that of the banks that had ex-ante similar likelihood to diversify but did not.

4.3.3 Propensity score matching

Definition of treatment, treated and control group

Treatment is defined as growth in interstate diversification over the period 2001-2006. (Geographic focus (standardized) in 2006 - Geographic focus (standardized) in 2001)/ (Geographic focus (standardized) in 2001)

Treatment group consists of those banks that in the pre-crisis period had a growth in interstate diversification (drop in focus) which was larger than 20%.

Control group consists of those banks that in the pre-crisis period had a growth in interstate diversification which was between +10% (increase in focus) and – 10% (increase in diversification).

Implementing the propensity score matching

These steps are followed in order to perform propensity score matching. First, propensity score for the treated and untreated banks is estimated; next, the treated are matched with the untreated and last, the average treatment effect on the treated is estimated.

Step 1: ***estimating the propensity score***
To estimate the propensity score *a logit regression* is used, with a dummy variable equal to one for treated banks (banks, which diversified significantly over the period preceding the crisis) and zero otherwise.

It is important to remind the reader that the main purpose of the propensity score estimation is not to predict the treatment, but to balance all the covariates between the two groups (Augurzky and Smidt, 2001). Hence, the vector of control variables should include all factors that affect both treatment and outcome. To do so, various bank-specific variables that reflect both general characteristics of bank financial statement structure (bank liquidity, loans and

deposits in asset structure) and the most frequently cited motivating factors for diversification are included in the analysis. It can be expected that banks decide for higher growth of geographic diversification because they want to improve their performance and reduce their risk. Banks' risk is measured with the level of Z-score and leverage, while the performance is measured with cost to income ratio (unfortunately, the analysis cannot include neither the net interest margin, because the net interest margin is highly correlated with loan loss provisions over loans, nor can the ROAE or ROAA, because they are highly correlated with cost to income ratio). In addition, those variables that control for the differences between banks in their change in risk are included in the model (please, see the previous section).

One of the required conditions in the propensity score matching analysis is that the variables included in the model should not be affected by the treatment. To ensure this, the bank specific variables employed in the propensity score model are taken for the year before the treatment, for the year 2000.

Step 2: *matching*

The nearest neighbor matching is applied and *a common support* is imposed. The imposition of common support means that treated observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls, are removed from the analysis. Furthermore, the nearest neighbor matching faces the risk of bad matches, in the case when the closest neighbor is far away. In order to avoid this problem a tolerance level for the maximum propensity score distance is imposed, so called *caliper* (or propensity range). The advantage of caliper is that bad matches are avoided and hence, the matching quality increases while the disadvantage of caliper is that, because fewer matches can be performed, the variance of the estimates increases. Applying caliper matching means that an individual from the comparison group is chosen as a matching partner for a treated individual that lies within the caliper and it is closest in terms of propensity score. A possible disadvantage of caliper matching is that it is difficult to know a priori what choice for the tolerance level is appropriate (Smith and Todd, 2005). Two values for caliper are used - 1% and 2%. *Replacement*[13] is allowed. Psmatch2 in Stata is applied. The results for the 2% value of caliper are discussed.

The sample used for the estimation of the average treatment effects is equal to 21 treated and their 13 controls. To verify the quality of matching, first the distribution of the propensity score for the treated and untreated before and after matching is plotted. As can be seen from Figure 3 in the matched sample, the propensity score distribution of the treated is very close to that of the matched sample.

[13] The results do not change significantly if replacement is not allowed. The main difference is the drop in the number of treated banks.

Given that matching is conditioned on the propensity score rather than on all covariates, it then needs to be checked whether the matching procedure is able to balance the distribution of all the relevant variables in both the control and treatment groups (Caliendo and Kopeinig, 2005). Rosenbaum and Rubin (1985) recommend using a two-sample t-test in order to check if there are statistically significant differences in covariate means between the treated and matched control groups. While differences in means between the treated and matched control groups before matching are expected, after matching these differences in covariates should not be statistically significant anymore. The results of the tests are reported in Table 10. As can be seen, there are some significant differences before matching, but in the matched sample the covariates are balanced in both groups, suggesting a successful matching.

More importantly, a test of balancing property of propensity score is performed. This test proposed by Becker and Ichino (2002), tests if observations with the same propensity score have the same distribution of characteristics independently of treatment status. It tests the balancing property for each variable in each block (it tests if means of each characteristic do not differ between treated and control units within each interval). The test which is not reported shows that the balancing property is satisfied.

Step 3: ***estimating the average treatment effect on the treated***
The matched sample is used to estimate the causal effects of growth in geographic diversification in the pre-crisis period on the indicator of bank performance, which is growth in risk during the crisis and is measured with an increase in Z-score over the period 2006-2009. The average treatment effect on the treated is estimated as the difference in the mean in the performance indicator between the treated and untreated banks.

The estimates of the average treatment effect on the treated are reported in Table 12. It can be said that the effect of growth in geographic diversification before the crisis on banks' risk is positive when the average treatment effects on the treated is positive, meaning that the increase in risk was smaller for the treated compared to untreated banks. In this paper a negative average treatment effects on the treated for banks that had a large growth in geographic diversification before the crisis is found. This means that those banks that diversified had a larger increase in risk during the crisis compared to those banks that did not.

5 Conclusion

The crisis which started in 2007 is the worst financial crisis since the Great Depression. One observation related to this period is a high variation in cross-section of banks' risk across US banks. Here lies the reason why this paper explores more in detail the clustering of banks with respect to pre-crisis financial statement information, before and after the crisis level of risk, and during the crisis change in risk. To further check these results, it next applies the regression estimation technique in order to analyze whether the relationship between banks'

geographic diversification and change in (and level of) risk during the crisis was also statistically significant. In the last part of the analysis, it applies matching to estimate causal treatment effects of an increase in banks' interstate diversification before the crisis on the change in banks' risk during the crisis.

This paper contributes to the existing literature in several ways. To begin with, the analysis applies the clustering technique, which is a very helpful tool in understanding the complex nature of multivariate relationships analyzed in this study. Moreover it is the first study that analyzes the effect of geographic diversification on banks' risk during the crisis. Other studies analyzing the crisis period focus on other issues (corporate governance issues, change in the composition of bank capital and dividend payments, bank lending, securitization), but none of them addresses the relationship analyzed in this paper. The issue of diversification is an important one because it is related to the optimum degree of diversification (the traditional banking theory suggests a well-diversified organization, while corporate finance theory suggests that a firm should stay focused), while the issue of banks' risk is important and relevant for several groups of stakeholders: for supervisors and regulators, managers of financial institutions and financial investors.

Middle-sized and large US banks - banks with total assets larger than 5 billion US dollars at the end of the year 2006 - are included in the analysis. A cluster analysis is performed in the first part of the analysis, regression analysis is conducted in the second part of the same analysis and the propensity score matching approach is applied in the last part of the analysis. Banks' risk is measured with Z-score while the geographic focus is measured with the Herfindahl-Hirschman index based on deposit dispersion.

Conclusions of the analysis are the following. Three groups of banks with respect to the pre-crisis financial statement information, before and after the crisis' level of risk, and during the crisis' change in risk can be observed. In general it holds that, if before the crisis, a bank had lower geographic diversification, lower profitability, higher Tier 1 ratio, lower ratio of mortgage loans over loans and lower level of Z-score, it also had a smaller increase in risk during the crisis. This result can be explained by the fact that most probably these banks were engaged in profitable business lines before the crisis and at the same time they underestimated the risk of this business - these banks had too low equity for the risk they took before the crisis. Results in the second part of the analysis indicate that the relationship between banks' geographic diversification and their change and level of risk during the crisis was not statistically significant. In the last part of the analysis, a negative average treatment effect on the treated for banks that had a large growth in geographic diversification before the crisis is found, which means that those banks that diversified significantly before the crisis had a larger increase in risk during the crisis compared to those banks that did not diversify much before the crisis.

These results have potential policy implications for the ongoing process of changing regulatory framework (Basel III). In particular these findings support the Basel Committee reform to strengthen capital rules (Basel Committee on Banking Supervision, 2010). First of all, these results indicate that the banks that were hit the most during the crisis (measured with an increase in risk during the crisis) were the ones which set up high leverage before the crisis and had a low quality of their capital base. Consequently, when the crisis hit, these banks had a capital buffer which was too low to cope with the crisis, so most of them received government capital injection in order to continue their operations successfully. Based on this finding it can be argued that the Basel Committee reform to raise the quality and quantity of the regulatory capital is an appropriate one and that the expected result of its imposition is that during the next crisis banks will be in a better position and they will need less government financial help to survive it. Second, Basel 3 also addresses procyclicality – banks will set up capital during an economic boom, which will then be available for banks to absorb losses during an economic slowdown and the crisis periods. Also this proposition finds support in the findings of this paper because those banks that were forced to increase capital during the crisis (banks from cluster 1) did not increase their lending activities during the crisis – their ratio of loans over total assets even decreased a bit over the period 2006-2009. Taking into consideration that supply of credit to the economy is a very important banks' function and given the evidence that those banks that were capital constrained during the crisis engaged less in lending activities, the author of this paper is of an opinion that this proposition is an appropriate one. The expected result of its imposition is that during the next crisis, banks will be better prepared for it and will provide better support (in the form of new lending activities) in restarting the economy and they will not deepen the recession by not providing new lending to the economy. Therefore, the discussion in this paragraph can be summed up by saying that results in this paper demonstrate the importance of capital buffer, especially when the crisis hits, implying that the regulatory aim to strengthen capital requirements and to introduce procyclcality is meaningful and justified given the evidence that emerged after the crisis.

References

Acharya, V.V., Gujral, I., Kulkarni, N. and Shin, H.S. (2011), Dividends and bank capital in the financial crisis of 2007-2009, unpublished working paper, National Bureau Of Economic Research.

Acharya, V.V., Gale, D. and Yorulmazer, T. (2008), Rollover risk and market freezes, unpublished working paper, Stern School of Business, New York University.

Acharya, V.V., Hasan, I. and Saunders, A. (2006) Should banks be diversified? Evidence from individual bank loan portfolios, *Journal of Business* **79**, 1355–1412.

Acharya, V.V. and Richardson, M. (2009) Causes of the financial crisis, *Critical Review* **21**, 195-210.

Acharya, V.V., Schnabl, P. and Suarez, G. (2009), Securitization without risk transfer, unpublished working paper, Stern School of Business, New York University.

ArcObjects (2002), Cluster Analysis Tool.
Augurzky, B. and Schmidt, C.M. (2001), The propensity score: A means to an end, unpublished working paper, IZA Discussion Paper.
Barajas, A., Chami, R., Cosimano, T. and Hakura, D. (2010), U.S. bank behavior in the wake of the 2007–2009 financial crisis, unpublished working paper, International Monetary Fund.
Barr, R. and Siems, T. (1994) Predicting bank failure using DEA to quantify management quality, unpublished working paper, Financial Industry Studies Working Paper.
Basel Committee on Banking Supervision (2010, rev June 2011) *Basel III: A global regulatory framework for more resilient banks and banking systems*, Bank for International Settlements, Basel.
Bayazitova, D. and Shivdasani, A. (2009), Assessing TARP, unpublished working paper, University of North Carolina.
Becker, S.O. and Ichino, A. (2002) Estimation of average treatment effects based on propensity scores, *The Stata Journal* **2**, 358–377.
Beltratti, A. and Stulz, R. (2009), Why did some banks perform better during the credit crisis? A cross-country study of the impact of governance and regulation, unpublished working paper, National Bureau Of Economic Research.
Berger, A.N. and Bouwman, C.H.S. (2009), Bank capital, survival, and performance around financial crises, unpublished working paper, University of South Carolina.
Berger, A.N. and Humphrey, D.B. (1992), Measurement and efficiency issues in commercial banking, unpublished working paper, Finance and Economics Discussion Series.
Bexley, JB., James, J.F. and Haberman, J. (2010), The financial crisis and its issues, unpublished working paper, Sam Houston State University.
Bologna, P. (2011), Is there a role for funding in explaining recent U.S. banks' failures?, unpublished working paper, International Monetary Fund.
Bureau of Economic Analysis, National economic accounts, http://www.bea.gov/national/index.htm
Bureau of Labor Statistics, Labor force statistics from the current population survey, http://data.bls.gov/pdq/SurveyOutputServlet?data_tool=latest_numbers&series_id=LNU04000000&years_option=all_years&periods_option=specific_periods&periods=Annual+Data
Bodnar, G.M., Tang, C. and Weintrop, J. (1998), Both sides of corporate diversification: The value impacts of geographic and industrial diversification, unpublished working paper, Wharton School, The University of Pennsylvania.
Boyd, J.H., De Nicolò, G. and Jalal, A.M. (2009), Bank competition, risk and asset allocations, unpublished working paper, International Monetary Fund.
Caliendo, M. and Kopeinig, S. (2008), Some practical guidance for the implementation of propensity score matching, unpublished working paper, IZZA Discussion Paper.
Camaraa, B., Lepetita, L. and Tarazia, A. (2010), Changes in capital and risk: An empirical study of European banks, unpublished working paper, University of Limoges.
Chari, V.V., Christiano, L.J. and Kehoe, P.J. (2008), Facts and myths about the financial crisis of 2008, unpublished working paper, Federal Reserve Bank of Minneapolis.

Cole, R.A. and White, L.J. (2010), Déjà vu all over again: The causes of the U.S. commercial bank failures this time around, unpublished working paper, DePaul University and Stern School of Business, New York University.
DeYoung, R. and Roland, K.P. (2001) Product mix and earnings volatility at commercial banks: Evidence from a degree of total leverage model, *Journal of Financial Intermediation* **10**, 54–84.
Demirguc-Kunt, A., Detragiache, E. and Merrouche1, O. (2010), Bank capital: Lessons from the financial crisis, unpublished working paper, International Monetary Fund.
Deng, S.E. and Elyasiani, E. (2008) Geographic diversification, bank holding company value, and risk, *Journal of Money, Credit and Banking* **40**, 1217–1238.
Elsas, R., Hackethal, A. and Holzhäuser, M. (2006), The anatomy of bank diversification, unpublished working paper, University of Munich.
Erkens, D., Hung, M. and Matos, P. (2009) Corporate governance in the 2007-2008 financial crisis: Evidence from financial institutions worldwide, unpublished working paper, ECGI - Finance Working Paper.
FED: Open Market Operations, http://www.federalreserve.gov/monetarypolicy/openmarket.htm. Retrieved 12th April 2011.
Goldsmith-Pinkham, P. and Yorulmazer, T. (2010) Liquidity, bank runs, and bailouts: Spillover effects during the Northern Rock episode, *Journal of Financial Services Research* **37**, 83-98.
Grosse, R. (2010). The global financial crisis: A Behavioral view, unpublished working paper, George Mason University.
Härdle, W. and Simar, L. (2007) *Applied Multivariate Statistical Analysis*, Springer, Berlin.
Heckman, J.J., Ichimura, H. and Todd, P.E. (1997) Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *The Review of Economic Studies* **64**, 605-654.
Huang, R. and Ratnovski, L. (2011) The dark side of bank wholesale funding, *Journal of Financial Intermediation* **20**, 248–263.
Igan, D. and Pinheiro, M. (2009), Exposure to real estate losses: Evidence from the U.S. banks, unpublished working paper, International Monetary Fund.
Ivashina, V. and Scharfstein, D. (2008), Bank lending during the financial crisis of 2008, unpublished working paper, Harvard Business School.
Kwan, S. and Eisenbeis, R. (1997) Bank risk, capitalisation and operating efficiency, *Journal of Financial Services Research* **12**, 117–31.
Laevena, L. and Levined, R. (2009) Bank governance, regulation and risk taking, *Journal of Financial Economics* **93**, 259-275.
Lindsey, L.B. (2007) Fear and greed: Why the American housing credit crisis is worse than you think, *The International Economy* **21**, 22–55, 68–73.
MBA (19th November 2009) Delinquencies continue to climb in latest MBA national delinquency survey, http://www.mbaa.org/NewsandMedia/PressCenter/71112.htm. Retrieved 12th April 2011.

Morgan, D.P. and Katherine, S. (2003), Geographic diversification in banking and its implication for bank portfolio choice and performance, unpublished working paper, Federal Reserve Bank of New York.
National Bureau Of Economic Research, Information on recessions and recoveries, the NBER business cycle dating committee, and related topics, http://www.nber.org/cycles/main.html. Retrieved 12th April 2011.
Reuters (February 29, 2009) Three top economists agree 2009 worst financial crisis since great depression; risks increase if right steps are not taken.
Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for casual effects, *Biometrica* **70**, 41-55.
Rotheli, T. F. (2010), Causes of the financial crisis: Risk misperception, policy mistakes, and banks' bounded rationality, *The Journal of Socio-Economics* **39**, 119–126.
Sarkisyan, A., Casu, B., Clare, A. and Thomas, S. (2009), Securitization and bank performance, unpublished working paper, Cass Business School, City University London.
Sharma, S. (1996) *Applied Multivariate Techniques*, Wiley, New York.
Smith, J.A.. and Todd, P.E. (2005) Does matching overcome LaLonde's critique of nonexperimental estimators?, *Journal of Econometrics* **125**, 305-353.
S&P: *S&P/Case-Shiller Home Price Indices | Americas,* www.homeprice.standardandpoors.com. Retrieved 12th April 2011.
Stiroh, K.J. (2004a) Do community banks benefit from diversification, *Journal of Financial Services Research* **25**, 135–160.
Stiroh, K.J. (2004b) Diversification in banking: Is noninterest income the answer?, *Journal of Money, Credit, and Banking* **36**, 853–882.
Stiroh, K.J. and Rumble, A. (2006) The darkside of diversification: The case of U.S. financial holding companies, *Journal of Banking and Finance* **30**, 2131–2161.
Taliaferro, R. (2009), How do banks use bailout money? Optimal capital structure, new equity, and the TARP, unpublished working paper, Harvard Business School.
Tarraf, H. (2010), Literature review on corporate governance and the recent financial crisis, unpublished working paper, Lawrence Technological University.
The New York Times (September 24, 2008) Transcript. President Bush's speech to the nation on the economic crisis.
Treasury (2011) About Financial Stability. http://www.treasury.gov/initiatives/financial-stability/about/Pages/default.aspx.
Veronesi, P. and Zingales, L. (2008) Paulson's gift, unpublished working paper, University of Chicago.
Wheelock, D.C. and Wilson, P.W. (1995) Explaining bank failures: Deposit insurance, regulation, and efficiency, *Review of Economics and Statistics* **77**, 689-700.

Figures

Figure 1: Dendrogram (dendogram is determined by using the balance sheet data for fiscal year 2006)

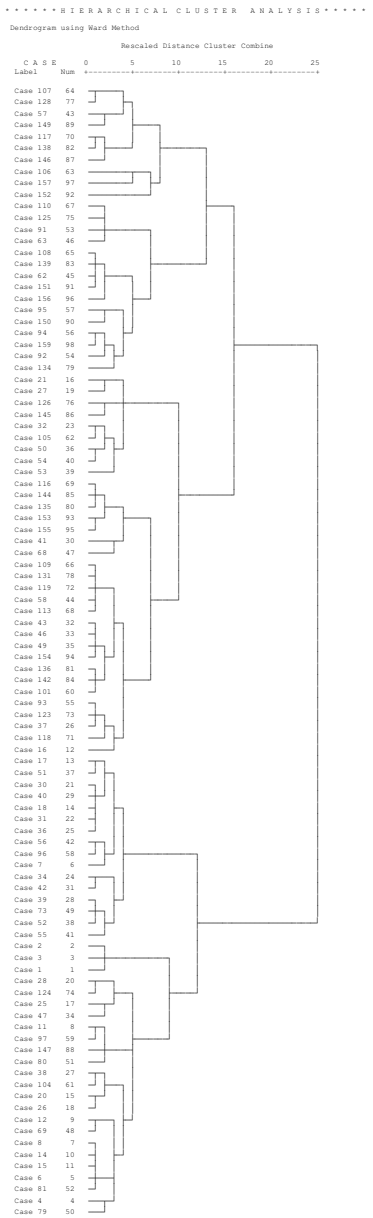
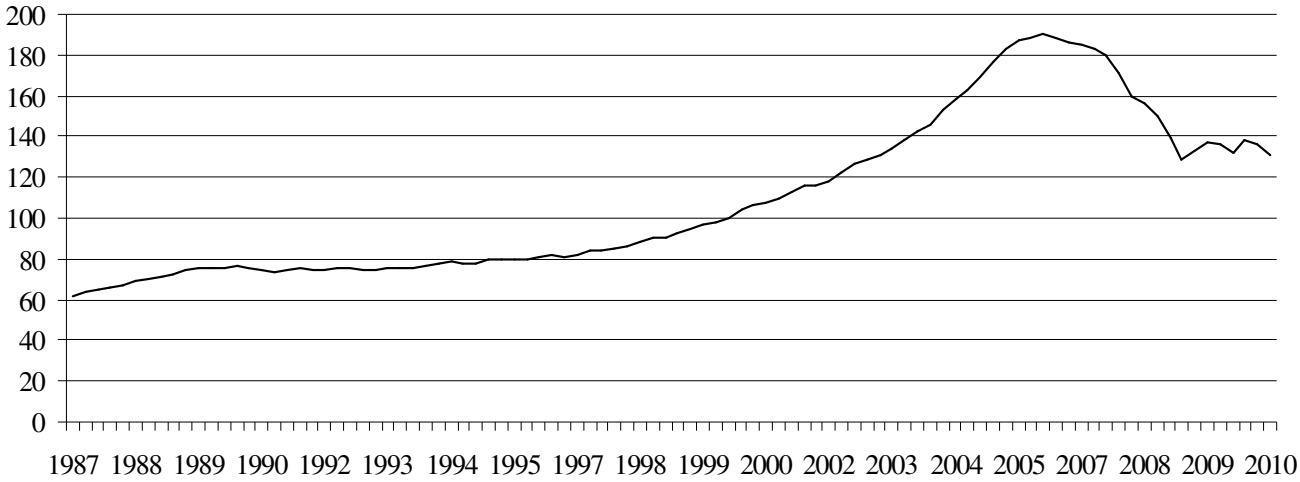


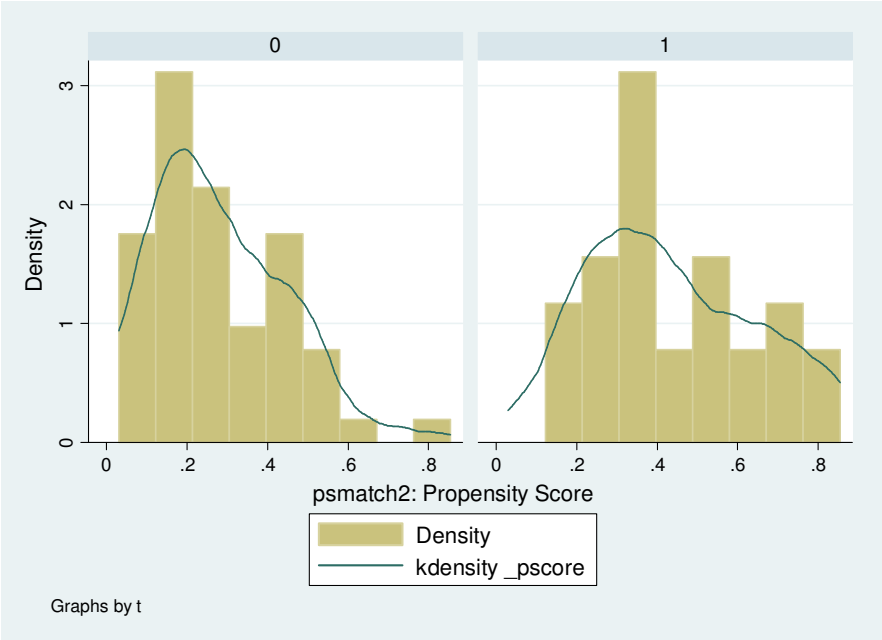
Figure 2: The S&P/Case-Shiller US National Home Price Index Level



Source: www.homeprice.standardandpoors.com.

Figure 3: Distribution of the propensity score of treated and untreated banks before matching and after matching

Unmatched Sample



Matched Sample

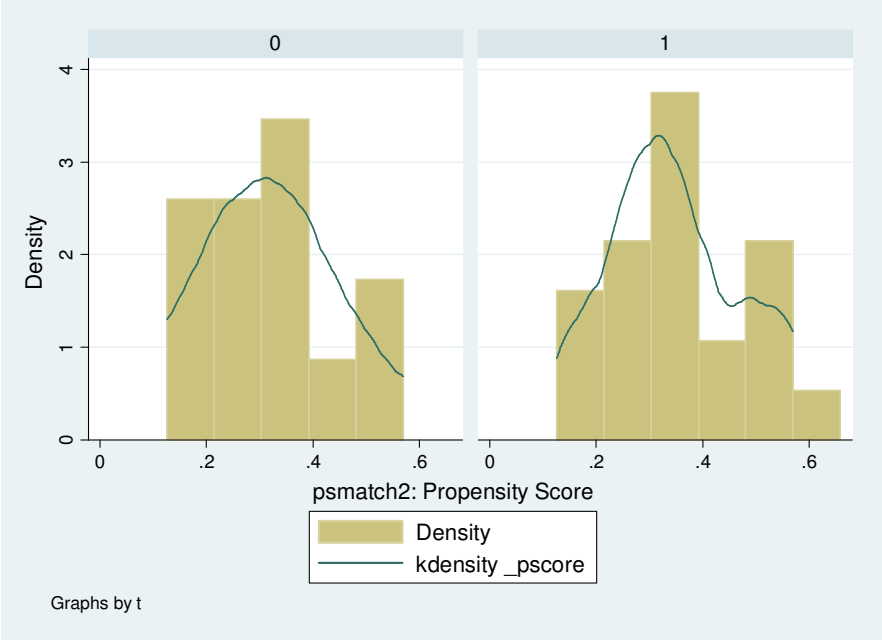


Table 1: Correlation coefficients between variables included in the cluster analysis (clusters are determined by using balance sheet data for the fiscal year 2006)

Variable	Geographic focus	Business focus	Ln_total assets	Cost-to-income	Equity_ta	Liquid assets_ta	Mortgage loans	Corporate loans	Deposits _ta	Ln_ $Zscore_{2009}$	Ln_ $Zscore_{2006}$
Business focus	0.448***	1									
Ln_total assets	-0.529***	-0.446***	1								
Cost-to-income	-0.130	-0.454***	-0.045	1							
Equity_ta	0.102	0.089	-0.055	-0.040	1						
Liquid assets_ta	-0.089	-0.201**	0.402***	-0.018	0.032	1					
Mortgage loans	0.103	0.158	-0.233**	-0.124	0.031	-0.168*	1				
Corporate loans	-0.249***	-0.243**	0.136	0.229**	-0.139	0.114	-0.309***	1			
Deposits_ta	0.006	0.091	-0.376***	0.144	-0.426***	-0.105	0.046	0.172**	1		
Ln_ $Zscore_{2009}$	0.108	-0.092	-0.249***	-0.001	0.233**	0.196**	0.191*	0.070	0.122	1	
Ln_ $Zscore_{2006}$	-0.137	-0.154	-0.028	-0.024	0.005	-0.127	0.186*	0.152	0.200**	0.029	1
Z-score $_{2009/2006}$	0.278***	0.158	-0.184*	-0.144	0.277***	0.337***	-0.051	-0.063	-0.100	0.559***	-0.511***

Notes: Variables are defined in section 4.1.1.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 2a: Mean of variables used in cluster analysis for years 2006 and 2009, clusters are determined by using the balance sheet data for year 2006

	Cluster 1	Cluster 1	Paird-Sample T-test	Cluster 2	Cluster 2	Paird-Sample T-test	Cluster 3	Cluster 3	Paird-Sample T-test	Total sample	Total sample	Paird-Sample T-test	ANOVA
Year	2006	2009		2006	2009		2006	2009		2006	2009		
N	40	40		33	33		25	25		98	98		
Geographic focus	0.470	0.435	2.56***	0.762	0.727	2.07**	0.889	0.846	1.31	0.675	0.640	3.27***	27.52***
Business focus	0.491	0.480	0.93	0.626	0.590	2.01**	0.672	0.604	1.96*	0.582	0.548	2.89***	15.19***
Interest revenue	*0.590*	*0.596*	*-0.43*	*0.746*	*0.707*	*2.21***	*0.761*	*0.713*	*1.65*	*0.686*	*0.663*	*1.91**	*12.82****
Trading revenue	*0.022*	*0.026*	*-1.04*	*0.005*	*0.007*	*-1.75**	*-0.003*	*0.020*	*-1.91**	*0.010*	*0.018*	*-2.30***	*2.42**
Comm & fee revenue	*0.205*	*0.198*	*0.56*	*0.085*	*0.088*	*-0.66*	*0.097*	*0.108*	*-0.58*	*0.137*	*0.138*	*-0.01*	*9.94****
Other revenue	*0.183*	*0.168*	*0.69*	*0.164*	*0.162*	*0.07*	*0.117*	*0.122*	*-0.04*	*0.160*	*0.155*	*0.49*	*2.06*
Ln_total assets	17.656	17.946	-4.85***	16.044	16.287	-8.55***	16.180	16.378	-4.21***	16.737	16.987	-8.69***	26.54***
Cost-to-income	0.623	0.737	-2.88***	0.567	0.656	-1.69*	0.564	0.609	-1.05	0.589	0.678	-3.37***	2.87*
Net interest margin	*0.035*	*0.033*	*2.02***	*0.035*	*0.034*	*1.34*	*0.032*	*0.032*	*-0.15*	*0.034*	*0.033*	*1.52*	*1.21*
ROAA	*0.013*	*-0.005*	*7.31****	*0.012*	*0.002*	*4.39****	*0.010*	*0.002*	*3.25****	*0.012*	*-0.001*	*8.65****	*3.58***
ROAE	*0.138*	*-0.040*	*7.88****	*0.126*	*0.036*	*4.93****	*0.105*	*0.007*	*2.85****	*0.126*	*-0.002*	*8.82****	*2.72**
Equity_ta	0.100	0.112	-1.97*	0.104	0.105	-0.19	0.111	0.114	-0.61	0.104	0.110	-1.76*	0.47
Tier 1	*0.094*	*0.108*	*-3.63****	*0.123*	*0.122*	*0.22*	*0.166*	*0.154*	*1.49*	*0.122*	*0.124*	*-0.52*	*11.02****
MM &ST funding_ta	*0.108*	*0.081*	*3.81****	*0.109*	*0.094*	*2.62****	*0.114*	*0.122*	*-0.50*	*0.110*	*0.096*	*2.44***	*0.03*
Deposits_ta	0.663	0.665	-0.18	0.716	0.714	0.30	0.679	0.672	0.42	0.685	0.683	0.25	1.55
Loans_deposits	*0.830*	*0.837*	*-0.437*	*0.785*	*0.808*	*-1.074*	*0.794*	*0.791*	*0.068*	*0.805*	*0.816*	*-0.793*	*0.338*
Loans_ta	*0.634*	*0.622*	*1.24*	*0.628*	*0.647*	*-1.48*	*0.599*	*0.607*	*-0.41*	*0.623*	*0.627*	*-0.44*	*0.44*
Mortgage loans	0.190	0.251	-2.20**	0.728	0.504	3.67***	0.063	0.116	-1.61	0.339	0.302	1.31	128.47***
Corporate loans	0.465	0.452	0.79	0.182	0.165	2.45**	0.170	0.196	-1.05	0.294	0.290	0.48	24.28***
Llprov_loans	*0.028*			*0.012*			*0.026*			*0.022*			*6.47****
Liquid assets_ta	0.074	0.077	-0.36	0.050	0.047	0.85	0.075	0.092	-1.20	0.066	0.071	-0.83	0.63
Ln_Zscore	2.662	0.743	9.56***	2.976	1.550	6.06***	2.020	1.233	2.15**	2.604	1.140	9.59***	10.86***
$Z\text{-score}_{2009/2006}$	-0.739			-0.546			0.030			-0.478			14.53***

Table 2b: Mean of variables used in cluster analysis for years 2006 and 2009, clusters are determined by using the balance sheet data for year 2009

	Cluster 1	Cluster 1	Paird-Sample T-test	Cluster 2	Cluster 2	Paird-Sample T-test	Cluster 3	Cluster 3	Paird-Sample T-test	Total sample	Total sample	Paird-Sample T-test	ANOVA
Year	2006	2009		2006	2009		2006	2009		2006	2009		
N	46	46		25	25		28	28		99	99		
Geographic focus	0.553	0.524	2.28**	0.777	0.761	1.21	0.804	0.733	2.18**	0.681	0.643	3.26***	7.75***
Business focus	0.511	0.472	2.54***	0.642	0.601	1.74*	0.644	0.630	0.64	0.581	0.549	2.91***	19.43***
Interest Revenue	*0.593*	*0.579*	*0.81*	*0.773*	*0.728*	*2.10***	*0.753*	*0.745*	*0.51*	*0.683*	*0.664*	*1.86**	*16.22****
Trading revenue	*0.016*	*0.024*	*-2.03***	*0.001*	*0.001*	*0.40*	*0.011*	*0.024*	*-1.32*	*0.011*	*0.018*	*-2.12***	*1.82*
Comm&fee revenue	*0.211*	*0.211*	*0.05*	*0.080*	*0.086*	*-0.60*	*0.089*	*0.085*	*0.47*	*0.144*	*0.144*	*0.08*	*17.45****
Other revenue	*0.179*	*0.160*	*0.92*	*0.145*	*0.141*	*0.12*	*0.122*	*0.137*	*-0.80*	*0.154*	*0.148*	*0.44*	*0.26*
Ln_total assets	17.518	17.816	-5.67***	15.910	16.111	-6.71***	16.250	16.500	-5.63***	16.753	17.013	-9.11***	21.96***
Cost-to-income	0.617	0.722	-3.00***	0.575	0.703	-1.90*	0.572	0.584	-0.44	0.594	0.678	-3.39***	2.55*
Net interest margin	*0.034*	*0.032*	*2.47***	*0.036*	*0.034*	*1.50*	*0.032*	*0.032*	*0.01*	*0.034*	*0.032*	*2.20***	*0.74*
ROAA	*0.013*	*-0.004*	*7.26****	*0.012*	*0.000*	*4.57****	*0.010*	*0.003*	*3.73****	*0.012*	*-0.001*	*8.97****	*2.62**
ROAE	*0.138*	*-0.041*	*7.30****	*0.130*	*0.018*	*5.10****	*0.102*	*0.035*	*4.07****	*0.125*	*-0.005*	*9.15****	*3.31***
Equity_ta	0.098	0.103	-1.50	0.102	0.102	0.02	0.108	0.116	-1.20	0.102	0.106	-1.57	2.26
Tier 1	*0.106*	*0.116*	*-2.32***	*0.125*	*0.125*	*0.02*	*0.137*	*0.134*	*0.42*	*0.120*	*0.123*	*-1.11*	*1.83*
MM&ST funding_ta	*0.105*	*0.072*	*4.81****	*0.103*	*0.078*	*3.80****	*0.115*	*0.136*	*-1.55*	*0.107*	*0.092*	*2.75****	*8.73****
Deposits_ta	0.687	0.694	-0.70	0.729	0.738	-1.31	0.676	0.646	2.48**	0.694	0.691	0.49	3.11**
Loans_deposits	*0.799*	*0.778*	*0.82*	*0.797*	*0.815*	*-0.61*	*0.788*	*0.815*	*-1.35*	*0.796*	*0.798*	*-0.13*	*0.39*
Loans_ta	*0.603*	*0.592*	*1.06*	*0.641*	*0.659*	*-1.11*	*0.606*	*0.623*	*-1.17*	*0.614*	*0.617*	*-0.51*	*1.80*
Mortgage loans	0.189	0.187	-0.16	0.693	0.701	-0.17	0.291	0.128	2.29**	0.348	0.300	1.45	66.80***
Corporate loans	0.438	0.433	0.52	0.168	0.146	1.31	0.220	0.230	-0.50	0.307	0.303	0.61	18.63***
Llprov_loans	*0.030*			*0.015*			*0.013*			*0.021*			*10.22****
Liquid assets_ta	0.084	0.102	-1.779*	0.039	0.036	0.59	0.072	0.062	1.930*	0.069	0.074	-0.92	3.95**
Ln_Zscore	2.634	0.658	9.32***	2.954	1.246	6.37***	2.180	1.811	2.06**	2.587	1.133	9.94***	9.39***
$Z\text{-score}_{2009/2006}$	-0.712			-0.652			-0.013			-0.499			17.38***

Table 2c: T test for mean difference in variables between three clusters, which are determined by using the balance sheet data for year 2006

	Cluster 1	Cluster 2	Mean Diff	T-test	Cluster 1	Cluster 3	Mean Diff	T-test	Cluster 2	Cluster 3	Mean Diff	T-test
Year	2006	2006			2006	2006			2006	2006		
N	40	33			40	25			33	25		
Geographic focus	0.470	0.762	-0.290	-4.88***	0.470	0.889	-0.419	-7.09***	0.762	0.889	-0.128	-2.22**
Business focus	0.491	0.626	-0.135	-4.23***	0.491	0.672	-0.181	-5.18***	0.626	0.672	-0.046	-1.17
Interest Revenue	*0.590*	*0.746*	*-0.156*	*-4.31****	*0.590*	*0.761*	*-0.171*	*-4.07****	*0.746*	*0.761*	*-0.014*	*-0.36*
Trading revenue	*0.022*	*0.005*	*0.017*	*1.44*	*0.022*	*-0.003*	*0.025*	*1.72**	*0.005*	*-0.003*	*0.007*	*1.09*
Comm & fee revenue	*0.205*	*0.085*	*0.119*	*4.19****	*0.205*	*0.097*	*0.108*	*2.86***	*0.085*	*0.097*	*-0.011*	*-0.42*
Other revenue	*0.183*	*0.164*	*0.020*	*0.74*	*0.183*	*0.117*	*0.066*	*1.92**	*0.164*	*0.117*	*0.046*	*1.27*
Ln_total assets	17.656	16.044	1.612	6.06***	17.656	16.180	1.476	4.88***	16.044	16.180	-0.136	-0.75
Cost-to-income	0.623	0.567	0.056	2.36**	0.623	0.564	0.059	2.03***	0.567	0.564	0.003	0.08
Net interest margin	*0.035*	*0.035*	*-0.000*	*-0.08*	*0.035*	*0.032*	*0.003*	*1.33*	*0.035*	*0.032*	*0.003*	*1.35*
ROAA	*0.013*	*0.012*	*0.002*	*1.53*	*0.013*	*0.010*	*0.003*	*2.33***	*0.012*	*0.010*	*0.002*	*1.71**
ROAE	*0.138*	*0.126*	*0.012*	*1.06*	*0.138*	*0.105*	*0.033*	*2.40***	*0.126*	*0.105*	*0.021*	*1.70**
Equity_ta	0.100	0.104	-0.003	-0.45	0.100	0.111	-0.011	-0.89	0.104	0.111	-0.007	-0.53
Tier 1	*0.094*	*0.123*	*-0.029*	*-3.35***	*0.094*	*0.166*	*-0.072*	*-4.45****	*0.123*	*0.166*	*-0.043*	*-2.07***
MM&ST funding_ta	*0.108*	*0.109*	*-0.001*	*-0.06*	*0.108*	*0.114*	*-0.006*	*-0.22*	*0.109*	*0.114*	*-0.005*	*-0.18*
Deposits_ta	0.663	0.716	-0.053	-2.18**	0.663	0.679	-0.016	-0.45	0.716	0.679	0.037	0.95
Loans_deposits	*0.830*	*0.785*	*0.045*	*1.01*	*0.830*	*0.794*	*0.036*	*0.52*	*0.785*	*0.794*	*-0.009*	*-0.12*
Loans_ta	*0.634*	*0.628*	*0.006*	*0.21*	*0.634*	*0.599*	*0.036*	*0.81*	*0.628*	*0.599*	*0.030*	*0.6731*
Mortgage loans	0.190	0.728	-0.538	-12.54***	0.190	0.063	0.127	2.91**	0.728	0.063	0.665	15.01***
Corporate loans	0.465	0.182	0.283	6.19***	0.465	0.170	0.295	5.16***	0.182	0.170	0.012	0.25
Llprov_loans	*0.028*	*0.012*	*0.016*	*4.44****	*0.028*	*0.026*	*0.002*	*0.33*	*0.012*	*0.026*	*-0.014*	*-2.59****
Liquid assets_ta	0.074	0.050	0.024	1.28	0.074	0.075	-0.001	-0.04	0.050	0.075	-0.025	-0.87
Ln_Zscore_{2006}	2.662	2.976	-0.314	-1.67*	2.662	2.020	0.642	3.53***	2.976	2.020	0.956	4.37***
Ln_Zscore_{2009}	0.743	1.550	-0.807	-3.41**	0.743	1.233	-0.490	-1.38	1.550	1.233	0.317	0.92
$Z\text{-}score_{2009/2006}$	-0.739	-0.546	-0.193	-2.31**	-0.739	0.030	-0.768	-4.85***	-0.546	0.030	-0.576	-3.09***

Table 2d: T test for mean difference in variables between three clusters, which are determined by using the balance sheet data for year 2009

	Cluster 1	Cluster 2	Mean Diff	T-test	Cluster 1	Cluster 3	Mean Diff	T-test	Cluster 2	Cluster 3	Mean Diff	T-test
Year	2009	2009			2009	2009			2009	2009		
N	46	25			46	28			25	28		
Geographic focus	0.524	0.761	-0.237	-3.21**	0.524	0.733	-0.209	-2.94**	0.761	0.733	0.028	0.44
Business focus	0.472	0.601	-0.129	-4.68***	0.472	0.630	-0.158	-6.08***	0.601	0.630	-0.029	-0.80
Interest Revenue	*0.579*	*0.728*	*-0.149*	*-4.38****	*0.579*	*0.745*	*-0.166*	*-4.90****	*0.728*	*0.745*	*-0.017*	*-0.47*
Trading revenue	*0.024*	*0.001*	*0.023*	*3.04****	*0.024*	*0.024*	*-0.000*	*-0.02*	*0.001*	*0.024*	*-0.023*	*-1.35*
Comm&fee revenue	*0.211*	*0.086*	*0.125*	*4.32****	*0.211*	*0.085*	*0.126*	*4.39****	*0.086*	*0.085*	*0.001*	*0.05*
Other revenue	*0.160*	*0.141*	*0.019*	*0.49*	*0.160*	*0.137*	*0.023*	*0.67*	*0.141*	*0.137*	*0.004*	*0.11*
Ln_total assets	17.816	16.111	1.705	5.37***	17.816	16.500	1.316	4.29***	16.111	16.500	-0.389	-2.25**
Cost-to-income	0.722	0.703	0.019	0.26	0.722	0.584	0.138	2.59**	0.703	0.584	0.119	1.58
Net interest margin	*0.032*	*0.034*	*-0.002*	*-1.12*	*0.032*	*0.032*	*0.000*	*0.13*	*0.034*	*0.032*	*0.002*	*1.18*
ROAA	*-0.004*	*0.000*	*-0.004*	*-1.18*	*-0.004*	*0.003*	*-0.007*	*-2.19***	*0.000*	*0.003*	*-0.003*	*-0.94*
ROAE	*-0.041*	*0.018*	*-0.059*	*-1.59*	*-0.041*	*0.035*	*-0.076*	*-2.31***	*0.018*	*0.035*	*-0.017*	*-0.62*
Equity_ta	0.103	0.102	0.002	0.33	0.103	0.116	-0.013	-1.76*	0.102	0.116	-0.015	-1.49
Tier 1	*0.116*	*0.125*	*-0.009*	*-1.05*	*0.116*	*0.134*	*-0.019*	*-1.79**	*0.125*	*0.134*	*-0.009*	*-0.77*
MM&ST funding_ta	*0.072*	*0.078*	*-0.006*	*-0.41*	*0.072*	*0.136*	*-0.063*	*-3.66****	*0.078*	*0.136*	*-0.058*	*-3.07***
Deposits_ta	0.687	0.729	-0.042	-1.64	0.687	0.676	0.011	0.36	0.729	0.676	0.053	1.67
Loans_deposits	*0.778*	*0.815*	*-0.037*	*-0.73*	*0.778*	*0.815*	*-0.037*	*-0.68*	*0.815*	*0.815*	*-0.000*	*-0.006*
Loans_ta	*0.592*	*0.659*	*-0.067*	*-1.86**	*0.592*	*0.623*	*-0.031*	*-0.81*	*0.659*	*0.623*	*0.036*	*1.15*
Mortgage loans	0.187	0.701	-0.514	-11.19***	0.187	0.128	0.060	1.21	0.701	0.128	0.573	9.59***
Corporate loans	0.433	0.146	0.287	5.66***	0.433	0.230	0.203	3.68***	0.146	0.230	-0.084	-1.94*
Llprov_loans	*0.030*	*0.015*	*0.015*	*3.02***	*0.030*	*0.013*	*0.018*	*3.65****	*0.015*	*0.013*	*0.003*	*0.87*
Liquid assets_ta	0.102	0.036	0.066	3.16**	0.102	0.062	0.040	1.49	0.036	0.062	-0.026	-0.98
Ln_Zscore_{2006}	2.634	2.954	-0.319	-1.58	2.634	2.180	0.454	2.38**	2.954	2.180	0.774	3.23**
Ln_Zscore_{2009}	0.658	1.246	-0.588	-1.98*	0.658	1.811	-1.152	-4.00***	1.246	1.811	-0.565	-2.38**
$Z\text{-}score_{2009/2006}$	-0.712	-0.652	-0.061	-0.72	-0.712	-0.013	-0.699	-5.23***	-0.652	-0.013	-0.639	-3.61***

Notes: Variables are defined in section 4.1.1.
***/**/* denotes statistical significance at the 1%/5%/10% level.
ANOVA: a one-way analysis of variance tests the hypothesis that several means are equal. In the Table 2a it is used information for 2006 and in the Table .2b it is used information for 2009.
In italic are reported those variables that are not used in cluster analysis, because of high correlation with other variables, but given that they may be interesting for a reader they are reported in the table.

Table 2e: Transition matrix

		ward_2009			Total
		1	2	3	
	1	35	1	3	39
ward_2006	2	3	22	8	33
	3	5	1	17	23
Total		43	24	28	95

Table 2f: Information about capital injection in capital purchase program (CPP)

Cluster (N)	Number of banks receiving CPP funds	Number of banks repaying CPP funds	Capital injection/Total assets*	Repaid capital /Received capital
1 (40)	29	19	2.35%**	72.06%
2 (33)	15	7	2.08%	56.11%
3 (25)	4	3	2.22%	75.00%
(98)	48	29	2.25%***	67.32%

Notes:
* Total assets refer to a year of received funds
** 2,42% if included also TIP
*** 2,30% if included also TIP

Table 3: Descriptive statistics for variables included in the first regression model

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
Z-score$_{2009/2006}$	-0.761	-0.547	0.589	-0.957	-0.366	111
Geographic focus	0.742	0.684	0.296	0.411	1.000	111
Business focus	0.391	0.443	0.210	0.290	0.584	111
Ln_total_assets	16.336	16.695	1.236	15.777	17.343	111
Equity_ta	0.091	0.107	0.074	0.077	0.108	111
Tier 1	0.104	0.129	0.104	0.091	0.123	108
Llprov_loans	0.017	0.026	0.024	0.008	0.035	111
Cost-to-income	0.606	0.615	0.229	0.561	0.664	111
Deposits_ta	0.699	0.672	0.147	0.619	0.757	111
Liquid assets_ta	0.038	0.066	0.097	0.024	0.062	111
Ln_Zscore	2.451	2.481	0.924	1.993	2.922	111

Notes: Variables are defined in section 4.1.1.

Table 4: Correlation coefficients between independent variables included in the first regression model

Variable	Geographic focus	Business focus	Ln_total _assets	Equity_ta	Tier 1	Llprov_ loans	Cost-to-income	Deposits_ta	Liquid assets_ta
Geographic focus	1								
Business focus	0.381***	1							
Ln_total _assets	-0.470***	-0.441***	1						
Equity_ta	0.044	0.212**	-0.069	1					
Tier 1	0.189**	0.342***	-0.174*	0.833***	1				
Llprov_loans	-0.049	-0.074	0.310***	0.038	0.018	1			
Cost-to-income	-0.017	-0.203***	-0.054	-0.060	-0.028	-0.014	1		
Deposits_ta	-0.025	0.026	-0.297***	-0.502***	-0.435***	-0.255***	-0.183**	1	
Liquid assets_ta	-0.053	-0.154	0.395***	0.149	0.261***	-0.053	0.010	-0.134	
Ln_Zscore	-0.152	-0.090	-0.057	0.088	-0.025	-0.132	-0.305***	0.282***	-0.113

Notes: Variables are defined in section 4.1.1. ***/**/* denotes statistical significance at the 1%/5%/10% level

Table 5: Regression estimation results: geographic focus and change in banks' risk during the crisis (OLS estimator)

Variables		(1)	(2)	(3)	(4)
Geographic focus	**b**	**0.223**	**0.126**	**0.155**	**0.067**
	se	0.170	0.167	0.148	0.152
Business focus	**b**	**-0.166**	**-0.221**	**-0.410**	**-0.486**
	se	0.341	0.387	0.333	0.376
Ln_total_assets	**b**	**-0.124***	**-0.144****	**-0.120***	**-0.145****
	se	0.065	0.073	0.063	0.073
Equity_ta	**b**	**1.482**		**2.565*****	
	se	1.050		0.999	
Tier 1	**b**		**0.533**		**0.985***
	se		0.634		0.562
Llprov_loans	**b**	**-3.208**	**-4.164***	**-4.085***	**-4.598***
	se	2.591	2.547	2.456	2.496
Cost-to-income	**b**	**-0.694*****	**-0.689*****	**-0.966*****	**-0.988*****
	se	0.209	0.220	0.248	0.251
Deposits_ta	**b**	**-0.445**	**-0.639**	**0.149**	**-0.287**
	se	0.602	0.730	0.578	0.708
Liquid assets_ta	**b**	**2.684*****	**2.783*****	**2.291*****	**2.365*****
	se	0.862	0.919	0.680	0.778
Ln_Zscore	**b**			**-0.258*****	**-0.234*****
	se			0.053	0.056
	Constant	**1.908**	**2.558**	**2.344**	**3.256***
	se	1.540	1.807	1.523	1.844
	F	4.016	4.582	5.786	6.052
	N	111	108	111	108
	r^2	0.379	0.379	0.498	0.484

Notes: Variables are defined in section 4.1.1. b refers to the regression coefficient, se to robust standard error. N is the number of observations.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 6: Descriptive statistics for variables included in the second regression model

Variable	Median	Mean	Std. Dev.	1st quartile	3rd quartile	N
Ln_Zscore_t	1.903	1.718	1.139	1.196	2.567	252
Geographic focus	0.689	0.647	0.296	0.393	0.968	252
Business focus	0.415	0.440	0.201	0.292	0.566	252
Ln_total_assets	16.443	16.859	1.271	15.934	17.480	252
Equity_ta	0.093	0.109	0.085	0.078	0.109	252
Tier 1	0.101	0.131	0.185	0.091	0.115	242
Llprov_loans	0.007	0.013	0.017	0.003	0.017	252
Cost-to-income	0.644	0.686	0.304	0.580	0.730	252
Deposits_ta	0.676	0.657	0.152	0.608	0.750	252
Liquid assets_ta	0.0350	0.0679	0.0952	0.0223	0.0653	252

Notes: Variables are defined in section 4.1.1.

Table 7: Correlation coefficients between independent variables included in the second regression model

Variable	Geographic focus	Business focus	Ln_total_assets	Equity_ta	Tier1	Llprov_loans	Cost-to-income	Deposits_ta
Geographic focus	1							
Business focus	0.281***	1						
Ln_total_assets	-0.490***	-0.397***	1					
Equity_ta	0.007	0.168***	-0.061	1				
Tier 1	0.065	0.175***	-0.057	0.828***	1			
Llprov_loans	0.009	-0.042	0.131**	0.031	0.037	1		
Cost-to-income	-0.011	-0.179***	-0.003	-0.100	-0.143**	0.089	1	
Deposits_ta	-0.008	0.000	-0.319***	-0.515***	-0.458***	-0.094	-0.116*	1
Liquid assets_ta	0.018	-0.196***	0.380***	0.110*	0.089	0.114*	0.015	-0.114*

Notes: Variables are defined in section 4.1.1.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 8: Regression estimation results: geographic focus and level of banks' risk during the crisis, over the period 2007-2008 (OLS estimator)

Variables		**(1)**	**(2)**
Geographic focus	**b**	**0.159**	**-0.015**
	se	0.190	0.218
Business focus	**b**	**-0.925*****	**-0.800****
	se	0.303	0.335
Ln_total_assets	**b**	**-0.032**	**-0.107**
	se	0.064	0.070
Equity_ta	**b**	**4.563*****	
	se	1.085	
Tier 1	**b**		**0.541**
	se		0.504
Llprov_loans	**b**	**-40.363*****	**-42.708*****
	se	3.613	4.061
Cost-to-income	**b**	**-1.221*****	**-1.376*****
	se	0.284	0.278
Deposits_ta	**b**	**1.413*****	**-0.387**
	se	0.528	0.671
Liquid assets_ta	**b**	**0.542**	**0.781**
	se	1.009	1.030
	constant	**2.536***	**5.586*****
	se	1.464	1.639
	F	32.562	31.456
	N	252	242
	r^2	0.670	0.637

Notes: Variables are defined in section 4.1.1. b refers to the regression coefficient, se to robust standard error. N is the number of observations.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 9: Determinants of treatment probability

Variables		Logistic regression
Geographic focus	**b**	**-1.156**
	se	1.167
Business focus	**b**	**-2.074**
	se	1.765
Ln_total_assets	**b**	**0.487***
	se	0.301
Liquid_ assets_ta	**b**	**-5.028**
	se	3.955
Loans_ta	**b**	**-1.687**
	se	2.139
Deposits_ta	**b**	**2.993**
	se	2.043
Equity_ta	**b**	**6.612**
	se	6.173
Llprov_loans	**b**	**8.783**
	se	34.725
Cost-to-income	**b**	**-0.222**
	se	0.753
Ln_Zscore	**b**	**-0.027**
	se	0.364
	constant	**-7.724**
	se	6.465
	N	84
	Pseudo r^2	0.137
	Log likelihood	-46.152

Notes: The dependent variable equals to one for treatment and zero for control observations. Variables are defined in section 4.1.1. Values for explanatory variables are taken for the fiscal year 2000.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 10: T-test for the equality of means of covariates before and after matching

	The unmatched sample			The matched sample		
Variable	**Control**	**Treated**	**Difference**	**Control**	**Treated**	**Difference**
Geographic focus	0.852	0.702	0.150***	0.797	0.799	-0.003
Business focus	0.480	0.411	0.069	0.531	0.432	0.100
Ln_total_assets	15.764	16.585	-0.821***	15.894	16.031	-0.137
Liquid_ assets_ta	0.078	0.072	0.006	0.085	0.063	0.022
Loans_ta	0.603	0.608	-0.004	0.596	0.633	-0.037
Deposits_ta	0.626	0.632	-0.006	0.719	0.671	0.048
Equity_ta	0.093	0.105	-0.012	0.094	0.085	0.009
Llprov_loans	0.006	0.008	-0.002	0.003	0.007	-0.004
Cost-to-income	0.694	0.622	0.072	0.550	0.620	-0.071
Ln_Zscore	2.335	2.623	-0.288	2.344	2.545	-0.201
N	56	28	84	13	21	34

Notes: Variables are defined in section 4.1.1.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 11: Correlation coefficients between independent variables

	Geographic focus	Business focus	Ln_total _assets	Liquid_ assets_ta	Loans_ta	Deposits_ta	Equity_ta	Llprov _loans	Cost-to-income
Business focus	0.233**	1.000							
Ln_total_assets	-0.447***	-0.512***	1.000						
Liquid_ assets_ta	0.006	-0.314***	0.406***	1.000					
Loans_ta	-0.201*	-0.147	-0.111	-0.505***	1.000				
Deposits_ta	-0.005	0.220*	-0.307***	-0.225**	0.305***	1.000			
Equity_ta	0.084	0.349***	-0.166	-0.062	-0.338***	-0.269***	1.000		
Llprov_loans	0.110	-0.115	0.093	0.001	0.008	-0.314***	0.477***	1.000	
Cost-to-income	0.096	0.192*	-0.307***	0.035	-0.257**	0.033	0.367***	0.085	1.000
Ln_Zscore	-0.309***	-0.063	0.128	-0.193*	0.163	0.137	0.061	0.107	-0.248**

Notes: Variables are defined in section 4.1.1.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 12a: Treatment effect on the treated (with replacement)

Caliper	Variable	Treated (N)	Controls (N)	Difference (se)
1%	Change in risk 2009-2006	-0.709	-0.455	-0.253*
		(18)	(13)	(0.156)
2%	Change in risk 2009-2006	-0.704	-0.440	-0.264*
		(21)	(13)	(-1.780)

Notes: se refers to bootstrapped standard error. N is the number of observations.
***/**/* denotes statistical significance at the 1%/5%/10% level.

Table 12b: Treatment effect on the treated (without replacement)

Caliper	Variable	Treated (N)	Controls (N)	Difference (se)
1%	Change in risk 2009-2006	-0.684	-0.356	-0.329**
		(13)	(13)	(0.169)
2%	Change in risk 2009-2006	-0.751	-0.383	-0.368**
		(15)	(15)	(0.155)

Notes: se refers to bootstrapped standard error. N is the number of observations.
***/**/* denotes statistical significance at the 1%/5%/10% level.