

RA **Economics and institutional change**

# Peer-Group Detection of Banks and Resilience to Distress

Andrea Flori  
Simone Giansante  
Fabio Pammolli

Research Area

**Economics and institutional change**

# Peer-Group Detection of Banks and Resilience to Distress

**Andrea Flori**

IMT School for Advanced Studies Lucca

**Simone Giansante**

School of management, University of Bath

**Fabio Pammolli**

Politecnico of Milan

ISSN 2279-6894

IMT LUCCA EIC WORKING PAPER SERIES #06/2016

© IMT School for Advanced Studies Lucca

Piazza San Ponziano 6, 55100 Lucca

# Peer-Group Detection of Banks and Resilience to Distress

Andrea Flori\*

Simone Giansante†

Fabio Pammolli‡

## Abstract

The paper looks at the importance of the true business model in shaping the risk profile of financial institutions. We adopt a novel indirect clustering approach to enrich the classic bank business model classification on a global data set including about 11,000 banks, both listed and non-listed representing more than 180 countries over the period 2005-2014. A comprehensive list of global distress events, which combines bankruptcies, liquidations, defaults, distressed mergers, and public bailouts, is regressed against financial statement ratios (i.e. proxies for CAMELS) and controlling for macro and sectoral effects using a rare-event logit model. Our findings suggest that individual characteristics along with macro and sectoral factors contribute differently, sometimes with opposite sign, to the likelihood of distress and to the volatility of business models with the exception of liquidity whose contribution appears exogenous to business model choice. By capturing the switching behaviour across groups, we find that business model volatility exacerbates vulnerability and distress, especially if moving from wholesale-oriented to deposit oriented groups.

## 1 Introduction

The financial instability of recent years has put banking activities under deep scrutiny. The contribution of banking business models to financial performance and resilience is a major concern for regulators aiming to monitor banking business models (BCBS, 2010; BCBS, 2011a). The recent Supervisory Review and Evaluation Process (SREP) put in place in Europe (EBA, 2014) is a direct example of business model analysis as an important pillar for monitoring security and stability of the banking sector. However, the variety of funding opportunities and investment strategies exacerbated by both financial innovation and the deregulation process of the pre-financial crisis period (Beltratti and Stultz, 2012; Diamond and Rajan, 2009) gave banks the potential strategic advantage of differentiating their activities to boost competitiveness and leverage their strength (Roengpitya et al., 2014). Hence, the constellation of possible combinations of asset-liability choices makes the classification of banking business models a complex task (Ayadi and De Groen, 2015; Bongini et al., 2015; Mergaerts and Vander Vennet, 2016). This study aims to address this issue by exposing the true business models adopted by banks as direct representation of their funding and investments strategies rather than relying on their official specialization. We claim that true business model identification plays an important role in shaping the risk profile of financial institutions by providing empirical evidence in the context of global distress events of financial institutions during the recent financial crisis.

The outbreak of financial markets of 2008 represents a unique opportunity to test the resilience of banking institutions as a function of their true activities. Due to the global magnitude of the crisis, our assessment exploits a large sample of financial institutions, which includes more than 11,000 banks and financial intermediaries, both listed and non listed as emphasized by Köhler (2015), representing more than 180 countries. A detailed list of balance sheet items collected from Bankscope and FDIC, along with macro-economic and sectoral indices retrieved from BIS, Datasream, OECD and WorldBank, is used to characterize and control banks' business models. A global

---

\*IMT School for Advanced Studies, Lucca: andrea.flori@imt.it

†University of Bath - School of Management: S.Giansante@bath.ac.uk (Corresponding author)

‡Politecnico of Milan: fabio.pammolli@polimi.it

distress even list is constructed by merging signals regarding bankruptcies and liquidations from Bankscope and FDIC, defaults from Moody’s and S&P, distressed mergers from both Bankscope, Moody’s and S&P and public bailouts from Laeven and Valencia (2012), all of them with global coverage.

Revealing the true banking business models is an ambitious task, especially if extended to a global perspective, where the high level of heterogeneity across countries, the multidimensionality of banks characteristics and the presence of missing values, make classification very challenging. To address this issue, we propose a novel approach in finance for business models classification able to disentangle the multi-dimensionality problem<sup>1</sup> of peer group assessment, which identifies groups of institutions adopting the same business model. We borrow a clustering approach well-established in the complex system field where *i*) similarities among pairs of financial institutions are measured by the cosine similarity of their vectors of attributes (see e.g. Dongen and Enright, 2012), representing a detailed list of balance sheet items (a recent application of this approach is shown in Puliga et al., 2016), and *ii*) the identification of business model is driven by a community detection algorithm (Blondel et al., 2008) based on the maximization of a modularity quantity (see e.g. Fortunato, 2010 and Newman and Girvan, 2004).

We find seven specific peer groups that enrich the three classic business model categories discussed in previous works (e.g. Ayadi et al., 2012; Demirgüç-Kunt and Huizinga, 2010; Köhler, 2015; Mergaerts and Vander Vennet, 2016; Roengpitya et al., 2014), such as the deposit-oriented, the wholesale-oriented and the investment-oriented models. The interpretation of these three traditional categories is mainly based on their funding characteristics. However, due to the mix features of the business models we detected, we label our identified peer groups according to the distinctive balance sheet characteristic that best differentiate each model from the others.

Our findings suggest that three business models are persistent over the whole period from 2005 to 2014, namely *Wholesale*, *Commercial* and *Saving* models. The first is a wholesale-oriented model characterised by dominant wholesale funding and diversified loans investments made up of mainly Russian and US bank holdings<sup>2</sup>. The second is another wholesale-oriented model represented by mainly European and US commercial banks and characterised by decent amount of wholesale funding and dominant commercial loans investments. The last one is a deposit-oriented business model dominated by US, Japanese and Indian institutions with mainly customer deposit funding and diversified commercial loans investments. As the result of a restructuring process during the financial crisis, we observe a *Diversified Retail* model with dominant deposit funding and large retail loans investments appearing since 2008. This deposit-oriented model is composed by many German and Swiss saving and cooperative banks that migrated from wholesale-oriented models at the peak of the financial crisis. In addition, we discover more unstable small groups emerging only for few years: a *Long Term* business model characterized by dominant long-term funding, commercial investments and composed by many European banks (mainly Italian and Spanish banks); a *Focus Retail* model with diversified funding and dominant retail loans investments made up of mainly Swiss saving banks and US bank holdings; an *Investment* model characterized by large non-interest income and non-deposit funding and dominated by US and UK international broker dealers. The first two groups are wholesale-oriented models that disappeared at the onset of financial markets in 2007, whereas the third is the only investment-oriented model we find, although emerging only after 2012. Note that different accounting standards across the sample and country-based regulations on bank specializations may have contributed to shape the structure of banks’ balance sheets. However, the normalization by total assets of the variables used to compute the cosine similarities is intended to mitigate this issue since accounting principles directly affect the size of balance sheet items (Roengpitya et al., 2014). In addition to that, the choice of exploiting a wide range of variables prevents the problem of a single balance sheet item (affected by accounting principles) impacting on the overall similarity. Indeed, the two main accounting standards, GAAP and IFRS, are well represented in all the three main business models and very stable overtime. This result is a further guarantee that our approach is not biased by different accounting standards.

<sup>1</sup>The multi-dimensionality issue refers to the variety of different assets and liabilities variables that can be used as individual comparison dimensions in peer group assessments. See e.g. Ayadi et al. (2011) and Mergaerts and Vander Vennet (2016).

<sup>2</sup>A detailed representation of the geographic distribution of financial institutions across business models is presented in Appendix D. The geographic map of each business model is discussed in terms of both number of institutions and their corresponding total assets.

The core of this study investigates the contribution of specific business models to the assessment of the risk of distress. We first confirm the soundness of our classification by performing a battery of non-parametric tests and multiple pairwise post-hoc comparisons. We show that our classification provides a better clustering identification than other methods adopted in literature, i.e. the wide-diffused Ward clustering algorithm (Ward and Joe, 1963), and the direct classification provided by Bankscope and used e.g. in Köhler (2015).

In order to capture distress events, we built a comprehensive list of global events by combining bankruptcies, liquidations, defaults, distressed mergers (see e.g. Betz et al., 2014; Vazquez and Federico, 2015), and public bailouts (Laeven and Valencia, 2010 and Laeven and Valencia, 2012). We regress these events against financial statement ratios (i.e. proxies for CAMELS) and controlling for macro and sectoral effects using a rare-event logit model (Firth, 1993) to take into account the possibility of a small amount of distress events. We confirm that CAMELS explain well the propensity of risk of the institutions, with high levels for ROA and liquidity reducing the risk of distress while ROE increases it. In line with the major findings in literature (Ayadi and De Groen, 2015; Demirgüç-Kunt and Huizinga, 2010; Köhler, 2015; Mergaerts and Vander Venet, 2016; Roengpitya et al., 2014), we observe that both wholesale-oriented and deposit-oriented models are prone to the risk of distress. However, we empirically observe different risk drivers contributing to that level of instability. First of all, the size of the institution, measured in terms of total assets, is positively correlated with the likelihood of distress only in the wholesale-oriented models, suggesting a higher vulnerability of large institutions adopting strategies with a much lower proportion of stable funding than the deposit-oriented group. Second, ROA impacts positively the risk of distress for deposit-oriented institutions, and negatively for the wholesale ones, suggesting that limited asset diversification of the deposit-oriented institutions would force them to concentrate their investments on fewer products in the pursuit of higher returns, with overall higher risk compared to the wholesale-oriented model. Opposite result for ROE suggesting that leverage may exacerbate stable funding in the case of deposit-oriented institutions, improving therefore their resilience, while wholesale-oriented institutions would pay the price of more unstable non-deposit funding growth in the pursuit of better returns on equity. The only exception we find is on liquidity whose impact on the risk of distress is negative and well consistent across business models, suggesting a contribution to resilience exogenous to business model choice (Tirole, 2011).

Another interesting finding is the impact of business model volatility at the institution level measured by the number of switches to a different business model prior to the financial crisis. We provide supporting evidence on a consistent portion of distresses coming from institutions that migrated quite often between peer groups prior to the crisis. More specifically, among the 204 distressed cases that we collected, almost 30% refers to institutions switching peer group at least once in the three years period before 2008. Those institutions with volatile business models were more likely to be in distress after the breakdown of financial markets. Our findings indicate that the higher the volatility of business model adopted, the higher the likelihood of distress.

Finally, we attempted to find the factors that are more likely to affect the likelihood of changes in the adoption of business models. We observe the presence of three core peer groups quite stable over time with a modest turnover, which confirms the existence of a certain stickiness in changing business model. However, our findings suggest that institutions that are more peripheral/dissimilar within their peer group tend to be more prone to change business model. On institution characteristics, equities contributes negatively to business model volatility capturing the higher shareholder control over managers in changing business direction as well as facilitating the search of funding in order to comply with the adopted model. Liquidity on the contrary provides better flexibility for managers to quickly adjust their strategy. Similar to the distress analysis, the impact of liquidity is consistent across business models, which echoes the previous findings of liquidity impacting banks regardless of the business model. We also find factors featuring opposite sign among business models. The short-term funding is an example that promotes volatility on wholesale-oriented institutions, while reinforcing business model stability on deposit-oriented ones because of the dominant contribution of deposits to short-term debt. At the peer group level, the *Commercial* model is characterised by higher probability of switching behaviour due to its similarities to both *Wholesale* and *Saving* groups on funding and investment strategy, respectively.

Drawing on our results, we elaborate policy recommendations. We find different risk drivers for

banks adopting competing business models<sup>3</sup>, discouraging the one-rule-fits-all approach in favour of a more appropriate targeted intervention coherent with the true banking business models.

The rest of the paper is organized as follows. Section 2 discusses the main literature on banking business models classification and contributions to performance and resilience. The main empirical hypothesis are outlined in Section 3. The data set constructed for the empirical analysis is described in Section 4. Section 5 introduces the methodology for the banking peer groups classification and presents the econometric setup. The discussion of the business models main features, geographical compositions and transitions is presented in Section 6. Section 7 provides a detailed risk assessment of the resilience of institutions to the distress events of 2008-10 and discusses the impact of their business model volatility prior to the crisis. Final conclusion and remarks are given in Section 8.

## 2 Literature Review

Our paper covers issues related to a large number of studies that analyse the classification criteria of banking business models and their impact on performance and resilience to the risk of distress. Many of them concentrate their effort on a small subset of institutions, mainly large listed banks of specific geographic locations, generally US and Europe with few exceptions.

The first set of contributions focuses on performance and risk by using banks characteristics as main explanatory variables. For instance, Demirgüç-Kunt and Huizinga (2010) investigates the mix of non traditional funding-investment models on a panel of 1,334 large global banks in 101 countries from 1995 to 2007. They show a non monotonic impact on risk where, starting from low level of nondeposit funding in combination with noninterest income activities, these non traditional banking strategies provide risk diversification benefits, whereas further increase in the levels of the mix enhances bank fragility. Note that even though risk diversification can be beneficial at the institution level to reduce their probability of default, diversification at a large scale could eventually increase the similarity among institutions, and therefore increasing the probability of joint failures with dangerous systemic consequences (Wagner, 2010). The so-called *dark side of diversification* can be avoided by securitization, offering non-linear diversification strategies via tranching portfolio investments (Oordt, 2013). DeYoung and Torna (2013) separate the non-interest income exposures of US depository institutions into different categories showing that, although they provide no evidence that those exposures affect the chances of failure of healthy institutions, only those non traditional activities that require banks to hold risky assets (investment banking, proprietary trading, venture capital, etc.) exacerbate the likelihood of already distressed institutions to fail. Their two steps empirical approach is designed to discriminate between troubled and failed US banks. Although the latter institutions are publicly reported by FDIC, the troubled ones had to be proxied by some sort of distress-ranking method. They also show that institutions holding risky non-interest income assets tend to be aggressive lenders, not well diversified (mainly risky loans portfolios) with low stable funding, suggesting a wholesale-funding strategy that resemble a core business model detected by our classification approach (the *Commercial* model, see Section 6.2). With particular emphasis on regulatory capital requirements and liquidity buffers, Vazquez and Federico (2015) perform a stability analysis on a large set of US and EU banks in the interval 2001-2009. Using leverage and non stable funding ratios as main explanatory variables, they claim that high leveraged banks with weak structural liquidity were more likely to default during the 2008-09 financial crisis. Their econometric analysis relies on a probit model of distress events constructed from banking activities information provided by Bankscope and enriched by data on systemic crisis (Laeven and Valencia, 2010), also included in our distress event list, and completed with macroeconomic and monetary variables. We note that they consider the entire set of mergers as distress events, while we circumscribe our list to a narrow definition of distressed mergers as also suggested by Betz et al. (2014). A sub-classification between domestic and cross-border global institutions also reveals that the first were more vulnerable to liquidity risk whereas the second were more exposed to solvency risks due to high leverage levels. The data sample constructed in our study is closely related to Vazquez and Federico (2015) in terms of number of institutions and countries coverage. The introduction of new accounting standards and the partial data availability

---

<sup>3</sup>In the paper we use interchangeably *business model* and *peer group* as the result of the clustering procedure. Due to the characteristics of the data set, we also use *banks* and *institutions* as synonymous.

from Bankscope up to 2004, which is discussed in Section 4, affect substantially the consistency of balance sheet characteristics across institutions. To deal with this issue, Vazquez and Federico (2015) propose for instance different configurations in their regression analysis. In this study we take those issues into account by considering the period 2005-14, which provides a much larger and consistent number of balance sheet features to deal with. Similar findings are reported by Beltratti and Stultz (2012) who considers a panel of the largest 500 global banks to assess the determinants of stock performance during the financial crisis. On top of the positive relations between exposure to US real estate and the likelihood of distress, banks with lower leverage, higher regulatory capital and deposits shown more resilience to distress compared with those high leveraged institutions financed on short-term money markets, where the latter were the institutions that performed very well before the collapse of the markets in 2007. Although this analysis covers the pre-crisis period and only a small sample of banks, they provide a detailed discussion on the resilience of banks' business models based on the mix of asset-liability structures in comparison with the universal banking model (Benston, 1994). Their contribution inspired the set up for our signal models. An interesting comparison of traditional vs nontraditional banking activities is presented by Lozano-Vivas and Pasiouras (2010). A panel of 752 listed global commercial banks is created to investigate banking activity impact on cost and profit efficiency estimated by a stochastic-frontier based model. Their study shows that the inclusion of nontraditional and noninterest activities increases cost efficiency whereas mix results are found for the profit efficiency, due to environmental and regulatory conditions. Lately importance has been given to those institutions recognised as systemically important banks (SIBs). An investigation of only systemically important global banks (SIGBs) is provided by Bongini et al. (2015) in which 70 of the world largest banks, out of which 28 were listed as systemically important by BCBS (2011b), are studied. Stock market reaction and potential distress is shown to have been driven by bank capital levels rather than its classification as SIGB, which supports our choice of extending this analysis to non SIGBs. Betz et al. (2014) focus more on the predictability<sup>4</sup> of banks vulnerability by designing an early warning model on the largest European banks. Using CAMELS rating system as main descriptive elements of banking business models, they develop a signal model on the distress events during the financial crisis by combining both direct bank failures (bankruptcies, liquidations and defaults) with indirect events like state support of distressed institutions and mergers in distress. They show that CAMELS are good indicators of bank distress and that macro-financial imbalances and sectoral indices of vulnerability improve the performance of model predictability<sup>5</sup>. Our paper is closely related to Betz et al. (2014) in terms of the use of CAMELS variables and the enlarged set of distress events. Our analysis can be interpreted as an extension of these works since it relies on a large data set of more than 11,000 global banks, both listed and non listed, from more than 180 countries with distress events which include bankruptcies, liquidations, defaults, distressed merges and public bailouts.

The second set of contributions focuses on the classification criteria of bank business models and their contribution to performance and risk. The mainstream approach to banking business models is either the direct one provided by the institutions themselves, also called qualitative approach (Köhler, 2015), or an indirect, quantitative, classification where hierarchical grouping algorithms are applied, most likely the Ward and Joe (1963) together with the stopping rule on the number of identified clusters based on Caliński and Harabasz (1974). Variables used to assess the similarities of business models are usually balance sheet data, i.e. dimensions over which banks are supposed to have a direct influence (Ayadi et al., 2011; Roengpitya et al., 2014). As income statements characteristics tend to reflect the interaction between the institution and the market, therefore not under the direct control of the institution itself, they are usually excluded from business model assessment and peer group identification<sup>6</sup>. The above classifications are then used as additional regressors to explain performance and risk as well as the likelihood of distress events. Köhler (2015) compares direct classification of listed vs non-listed banks on a panel of 3362 European institutions in the period 2002-2011. Listed banks tend to show more investment type model, non-traditional activities with good proportions of nondeposit funding and noninterest income products, while unlisted banks are mainly saving and cooperative banks with more retail-

<sup>4</sup>A detailed review of banking failure prediction models is given in Demyanyk and Hasan (2010).

<sup>5</sup>Cole and White (2012) provide one of the first attempts to study US banks' failures using CAMELS proxies.

<sup>6</sup>An exception to this rule is presented in Puliga et al. (2016).



based models. Köhler (2015) confirm results shown in Demirgüç-Kunt and Huizinga (2010) on the non monotonic effect of non traditional banking activity to risk. This study motivates our choice to extend the sample to all possible global institutions, both listed and nonlisted. An indirect classification of European banks business model is provided by Ayadi and De Groen (2015). They use balance sheet data on 2,528 EU banks from 2005 to 2014 to classify them into five main models, namely wholesale, focused retail, diversified retail (type I and II) and investment. Whereas retail-oriented models are characterized by the classical customer deposit-loan intermediation activity (with different levels of diversification among the three groups), the wholesale and investment models depict more non traditional banking activities with heavy reliance on interbank lending and funding (wholesale), trading assets and derivatives (investment). More retail-oriented models show higher distance to default (z-score used to assess risk), in particular focused retail and diversified retail type I. Performance is also in favour of focused retail banks, with the highest performance among the groups. Our model classification resembles the same three model categories found above, with very similar individual characteristics that we extend on a global basis with a much larger sample (see Section 6). On a narrower list of balance sheets variables, Roengpitya et al. (2014) implements the same indirect approach as Ayadi et al. (2011), Ayadi et al. (2012) and Ayadi and De Groen (2015) on 222 international banks, along with some subjective judgemental element to filter the final groups. They found three major models used globally: retail-funded, wholesale-funded and trading. Retail and wholesale funded models are found to perform better during the period 2006-2014 than trading model, while retail banks show more volatility in their performance than wholesale institutions. They also analyse the transition of banking models over time, observing switches of banks from a retail-oriented towards a wholesale type in the 2005-07 period. The transition reverts in the period 2008-13, while trading banks look very stable over time). We also perform an inter-temporal analysis by employing a switch analysis to study the likelihood of institutions to migrate from a business model group to another. Due to the larger institution sample, our findings complete and expand those reported in the above study. Alternative approaches to business model classifications are presented in Mergaerts and Vander Vennet (2016). They propose a continuous type classification method to discover business models by employing a factor analysis among a set of large 505 EU banks in the period 1998-2013. They find two main factors describing the continuous factor loading ranges of asset, liability, income and capital structure of the banks. The first factor represents a retail-based business model, while the second represents a diversified model. They also find higher long-term profitability and high resilience of retail-based models compared with the diversified one. However, factor analysis only provides a description of the business models that better explain the variables set without a direct link with the individual institutions adopting that model. This is an important limitation as it would make any direct intervention by the regulators a complex task. The approach proposed in this paper preserves the individual bank classification feature guaranteed by indirect methods. It also advances on the accuracy of model specifications (strong point in the factor analysis) by employing the state-of-the-art set of individual bank characteristics on a large sample. This could be achieved by a clustering approach, which is the *Louvain* method (Blondel et al., 2008), that is specifically designed for large, sparse and complex data sets. Comparisons between our approach and the above methods are discussed in Section 6

In recent years, a growing literature has been devoted to studying leverage dynamics, its cyclically with asset growth and financial cycle, and active balance sheet management as a direct consequence of the business model adopted by banks. Although the issue is not directly related to resilience and distress, the literature in this area provides a useful framework to interpret the balance sheet dynamics across different business models in the run up of the crisis. Adrian and Shin (2010) propose a leverage adjustment model among top US broker dealers, characterised by a large proportion of wholesale debt. They claim that during economic upturn, the increase of asset prices translates into a positive boost of equity capital in the balance sheet of the banks. If banks do not react (passive management), the total asset rise would have a negative impact on leverage. However, when actively managed, the bank can expand its assets side in search for new investments that are financed by short-term debt (facilitated by higher capital). This mechanism, in conjunction with high markets liquidity, promote asset growth and leverage (see e.g. Halling et al., 2016). An opposite leverage adjustment is observed during economic downturns. Their so-called *feedback effect* resemble the *spiral effect* of Brunnermeier and Pedersen (2009). The impact



of wholesale funding to leverage pro cyclicalities is made explicit by Damar et al. (2013) who categorised Canadian financial institutions by their level of wholesale debt. They show that wholesale banks display higher degree of leverage procyclicality due to the readily available funds compared to deposit-oriented business models when actively managing their balance sheets. This effect is also exacerbated by highly liquid markets during economic booms. On the other side, short-term wholesale funding markets are also associated with higher risk of disruption during financial crisis than the deposit funding market, usually covered by deposit insurance. The discussion of our findings in relation with this literature can be found in Section 7.

### 3 Main Hypotheses

The research above suggests that the mix funding/investment strategy adopted by banks do contribute to their resilience, although they feed into the likelihood of distress in different, sometimes non linear, ways. Our first test aims at validating our classification approach, the key feature of our main hypothesis on the likelihood of distress. We therefore test the following hypothesis:

**Hypothesis H0.** The *Louvain* method enriches the current state-of-the-art banks classification providing superior clustering outcomes.

It is well established in the literature above that wholesale-funded institutions are more likely to face distress during economic downturn. We are interested in how individual characteristics, playing the role of risk drivers, contribute to the likelihood of distress as a function of the business model they adopt. By controlling for macro and sectoral effects, we test the following hypothesis:

**Hypothesis H1.** The impact of individual risk drivers to distress is a function of the business model.

Furthermore, the individual business model characteristic is time dependent and institutions do sometimes switch to different peer groups. As we find evidence of a substantial proportion of distress banks showing volatile business models prior to the crisis, we are interested in testing the following hypothesis:

**Hypothesis H2.** The higher the volatility of a bank to change business model, the higher the probability of that bank to experience distress.

To conclude our understanding of business models dynamics, we will also attempt to explore main driving forces that contribute to the transition of institutions to different business models prior to the crisis. Following the recent debate on liquidity and leverage, along with the similarity of each institution with their peers and controlling for macro and sectoral effects, we test the following hypothesis:

**Hypothesis H2bis.** The impact of individual characteristics to business model volatility is a function of the business model.

### 4 Data

We chose the interval 2005-14 as reference period to avoid the effects of new accounting standards IAS-IFRS introduced in the early 2000s<sup>7</sup>. In addition to that, the coverage of institutions in Bankscope improves substantially in 2005 with the inclusion of many European banks<sup>8</sup>. To investigate the presence of similarities across institutions in different countries we consider a global

<sup>7</sup>For a timeline of the relevant changes in IAS-IFRS, the interested reader can refer to FASB (2016).

<sup>8</sup>For some countries the number of institutions that are present increases significantly from 2004 to 2005. For instance, this is the case of: Italy (from 49 to 674), Spain (from 58 to 165), Germany (from 1472 to 1757) and Norway (from 50 to 104). The coverage of European institutions increases by about 40% from 2004 to 2005.

sample covering more than 180 countries (more than 100 countries with at least 10 institutions). Finally, our classification procedure is intended to overcome some of the potential limits due to broad classifications. For this reason we include all types of peer groups reported by Bankscope with the exception of Central Banks and Specialized Governmental Credit Institutions. Treatment of the data follows suggestions reported in Duprey and Lé (2014).

## 4.1 Features used to Characterize Business Models

The procedure used to identify peer groups relies on the characterization of financial institutions by means of balance sheet items retrieved from Bankscope. The choice to discard income statement variables is in line with previous literature (e.g. Ayadi et al., 2011; Roengpitya et al., 2014) and reflects the need to identify economic dimensions over which financial institutions can have a direct influence. In our study we exploit a detailed set of balance sheet items for both assets and liabilities, providing a very granular representation of banking activities. Table 1 reports summary statistics of the balance sheet variables (standardized by the total assets of the respective institutions) used to compute the similarities across financial institutions<sup>9</sup>. Annual mean values suggest that on average these balance sheet items do not have great variations over the reference period, except variables referred to mortgage, loans and deposits which experienced more volatile figures. Interestingly, we notice that within each variable there is a quite consistent variability (column *Pooled Std* in Table 1) and equity seems to have one of the most dispersed distribution. To avoid distortions due to the presence of institutions with total assets ranging among different levels of magnitude, we express each measure in the vector as a percentage of the respective total assets<sup>10</sup>.

**Table 1: Balance Sheet Variables Statistics.** Column *Annual Average* refers to the average across annual mean values from 2005 to 2014; column *Std (Annual Average)* shows the standard deviation of annual mean values; column *Pool Average* indicates the average when observations are pooled across the entire interval 2005-14; column *Pool Std* stands for the standard deviation of pooled data; column *Annual Average NAs* exhibits the average number of missing values computed across annual mean values. The average number of institutions per year is about 10400. Source: Bankscope, Authors' own elaborations.

Balance Sheet Items	Annual Average	Std (Annual Average)	Pool Average	Pool Std	Annual Average NAs
At-Equity Investments in Associates	0.01	0.00	0.01	0.05	3211.60
Available for Sale Securities	0.11	0.02	0.11	0.13	4563.20
Cash and Due from Banks	0.05	0.01	0.05	0.09	218.30
Corporate & Commercial Loans	0.24	0.03	0.23	0.25	4688.80
Customer Deposits (current)	0.29	0.02	0.29	0.24	2548.20
Customer Deposits (savings)	0.22	0.02	0.22	0.19	5895.00
Customer Deposits (term)	0.28	0.02	0.28	0.25	2551.30
Deposits from Banks	0.13	0.01	0.13	0.17	2579.90
Derivatives	0.02	0.00	0.01	0.05	7004.50
Fixed Assets	0.02	0.00	0.02	0.04	125.40
Held to Maturity Securities	0.06	0.00	0.06	0.09	6975.30
Loans and Advances to Banks	0.15	0.01	0.15	0.17	952.30
Other Assets	0.03	0.00	0.03	0.08	16.10
Other Consumer/Retail Loans	0.14	0.01	0.14	0.21	7073.20
Other Deposits and Short-Term Borrowings	0.11	0.02	0.10	0.17	4854.10
Other Funding	0.04	0.00	0.04	0.11	7798.10
Other Liabilities	0.04	0.01	0.04	0.50	625.80
Other Loans	0.38	0.03	0.37	0.32	1428.50
Other Mortgage Loans	0.09	0.05	0.12	0.16	9816.60
Other Securities	0.14	0.01	0.14	0.15	4641.70
Repos and Cash Collateral	0.07	0.01	0.07	0.14	9295.20
Reserves for Impaired Loans/NPLs	0.03	0.00	0.03	0.18	2819.50
Reserves for Pensions and Other	0.01	0.00	0.01	0.02	3765.70
Residential Mortgage Loans	0.29	0.06	0.27	0.25	7453.40
Reverse Repos and Cash Collateral	0.06	0.02	0.05	0.12	9229.50
Senior Debt Maturing After 1 Year	0.15	0.01	0.15	0.22	5749.00
Subordinated Borrowing	0.01	0.00	0.01	0.04	5901.00
Total Equity	0.12	0.02	0.12	0.59	607.30
Trading Securities and at FV through Income	0.05	0.01	0.04	0.11	4829.70

<sup>9</sup>The peer group detection is done for each year from 2005 to 2014 separately. From the set of institutions to be classified each year we exclude institutions for which we have less than 1/3 of the selected balance sheet items.

<sup>10</sup>Roengpitya et al. (2014) argue that this choice of using ratios is also useful to prevent distortions due to not homogeneous accounting standards under different regulations.

## 4.2 Distress Events

Table 2 shows the distress events definitions and the corresponding sources. As emphasised in the Introduction, institutions may be under distressed conditions due to several reasons, although the recent financial crisis suggests that government bailouts and state aid had a role in the avoidance of systemic crisis and cascade of banks failures. Therefore, direct failures are quite rare and presenting estimates separately for different distress events would have made the econometric estimation not robust enough. For these reasons, we propose a comprehensive set of distress events which take into account several definitions of bank distress (for an approach similar to ours see e.g. Betz et al. (2014) and Vazquez and Federico (2015), while Kick and Koetter (2007) distinguish between different types of distress).

**Table 2: Distress Events Definitions and Sources.** The amount of distress events refers to the period 2008-10. Distress categories are not mutually exclusive. The proportion of distressed institutions in the risk assessment is 204/8526.

Distress Event	Description	Source	Freq.
<b>Bankruptcy</b>	<i>This event occurs if the net worth of the bank falls below a country specific regulatory threshold</i>	<i>Bankscope</i>	<b>8</b>
<b>In Liquidation</b>	<i>This event is related to the sale of the bank by the liquidator as per the guidelines of the country regulations</i>	<i>Bankscope</i>	<b>17</b>
<b>Default</b>	<i>This event occurs when bank failed to repay interests or principal on its financial obligations beyond any grace period or if some of its instruments are replaced by other obligations at a diminished value as a consequence of a distressed exchange between counterparts</i>		<b>26</b>
	<ul style="list-style-type: none"> <li>• Short-Term rating NP (Not Prime) and Long-Term rating Caa (speculative or poor standing) or below</li> <li>• Short-Term rating SD (Selective Default) or D (Default) and Long-Term rating CCC (vulnerable) or below</li> </ul>	<i>Moody's</i> <i>Standard &amp; Poor's</i>	
<b>Distressed Merger</b>	<i>Merger (from Bankscope) which involves a bank with distressed conditions</i>		
	<i>Coverage Ratio in <math>t-1 &lt; 0</math></i> <ul style="list-style-type: none"> <li>• <math>(\text{Equity} + \text{Reserves for NPLs} - \text{Total Impaired Loans}) / \text{Total Assets} &lt; 0</math></li> </ul>	<i>Bankscope</i>	<b>19</b>
	<i>Short or long ratings below vulnerable state in <math>t</math></i> <ul style="list-style-type: none"> <li>• Short-Term rating NP (Not Prime) or Long-Term rating Caa (speculative or poor standing) or below</li> <li>• Short-Term rating SD (Selective Default) or D (Default) or Long-Term rating CCC (vulnerable) or below</li> </ul>	<i>Bankscope/Moody's</i> <i>Bankscope/S&amp;P</i>	<b>3</b>
<b>Public Bailout</b>	<i>State Aid from Governments, such as: Nationalization, Recapitalization and Assets Guarantees and Purchases</i>	<i>Laeven and Valencia (IMF, 2010-2012)</i>	<b>77</b>
<b>Other</b>	<i>US Failed bank list from the Federal Deposit Insurance Corporation</i>	<i>FDIC</i>	<b>63</b>

Bankruptcy occurs if the net worth of the bank falls below a country-specific regulatory threshold, while liquidations concern the sale of the bank by the liquidator as per the guidelines of the country regulations and the distribution of its assets to claimants. These two distress events are quite rare during the financial crisis as governments interventions created a safe net to prevent cascade failures. We therefore introduce additional distress definitions to capture these interventions. Defaults occurs if the bank failed to repay interests or principal on its financial obligations beyond any grace period or if some of its instruments are replaced by other obligations at a diminished value as a consequence of a distressed exchange between counterparts. We rely on ratings from Moody's and Standard & Poor's to assess the presence of a default state. In particular, we merge both short-term and long-term ratings and only if the evaluation of bank conditions is poor in both cases we consider that bank under a default event. Moreover, forced mergers of distressed institutions have occurred during the crisis. We define an institution as part of a distressed merger if its coverage ratio in  $t-1$  was negative. Coverage ratio is a typical indicator used to assess banks vulnerability conditions (see e.g. González-Hermosillo, 1999) and is computed as the sum of equity plus reserves for non-performing loans minus total impaired loans over total assets<sup>11</sup>. In addition, we consider as distressed mergers those cases where the institutions present a rating indicating

<sup>11</sup>For the impact of all mergers, not just distressed ones, to distress, see Vazquez and Federico (2015).

a vulnerable state. Finally, we enrich the data set of distressed institutions by including the information of public bailouts (Laeven and Valencia, 2010; Laeven and Valencia, 2012) and, for the US perimeter, we integrate Bankscope with bankruptcy information from the Federal Deposit Insurance Corporation<sup>12</sup>.

### 4.3 Indicators of Vulnerability

**Table 3: Measures, Data Description and Sources.** This table shows for each regressor its definition and the source from which we retrieved data. In some specifications of the models we consider aggregated regional proxies for macro and sector measures according to World Bank geographical classifications. Data for House Prices and Credit to Non-Financial Sector are on a quarterly basis and then annualized. Data from Datastream are daily and then annualized.

Measures	Description	Source
Capital	<i>Equity/Total Assets</i>	<i>Bankscope</i>
Capital Funding Ratio	<i>(Common Equity + Subordinated Borrowing)/Total Assets</i>	<i>Bankscope</i>
Roa	<i>Return on Average Assets</i>	<i>Bankscope</i>
Roe	<i>Return on Average Equity</i>	<i>Bankscope</i>
Cost to Income Ratio	<i>Operating Expenses/Operating Income</i>	<i>Bankscope</i>
Net Interest Margin	<i>Net Interest Revenues/Total Earning Assets</i>	<i>Bankscope</i>
Interest Expenses to Total Liabilities	<i>Total Interest Expenses/ Total Liabilities</i>	<i>Bankscope</i>
Liquid Assets to Short - Term Funding	<i>(Cash and other liquid assets)/(Deposits and ST Funding)</i>	<i>Bankscope</i>
Deposits to Total Funding	<i>(Deposits and ST Funding)/Total Funding</i>	<i>Bankscope</i>
Total Securities to Total Assets	<i>Total Securities/Total Assets</i>	<i>Bankscope</i>
GDP per capita	<i>Annual percentage growth rate of GDP per capita</i>	<i>World Bank</i>
Inflation	<i>Consumer prices (annual %)</i>	<i>World Bank</i>
House Price	<i>Real house prices</i>	<i>OECD/BIS</i>
Unemployment	<i>Share of the labor force that is without work but available for and seeking employment</i>	<i>World Bank</i>
FDI-Inflows	<i>Net inflows (% of GDP)</i>	<i>World Bank</i>
FDI-Outflows	<i>Net outflows (% of GDP)</i>	<i>World Bank</i>
Central Gvt. Debt	<i>Entire stock of direct government fixed-term contractual obligations (% of GDP)</i>	<i>World Bank</i>
Gvt. Long-Term Yield	<i>Long-term government bond yield</i>	<i>Datastream</i>
Bank NPLs to Gross Loans	<i>Nonperforming loans divided by the total value of the loan portfolio</i>	<i>World Bank</i>
Credit to Non-Financial Sector	<i>Banks domestic credit to non financial sector</i>	<i>BIS</i>
Market Index	<i>S&amp;P Global Equity Indices (annual % change) or, alternatively, FTSE Indices</i>	<i>World Bank/Datastream</i>
Sector Index	<i>FTSE Financial Indices</i>	<i>Datastream</i>
Stock Traded	<i>Number of traded shares multiplied by their respective matching prices</i>	<i>World Bank</i>

We use a wide set of regressors retrieved from several sources to gauge banks probability of distress. As suggested in literature (see e.g. Flannery, 1998; González-Hermosillo, 1999), bank-specific features can be expressed in terms of proxies for CAMELS variables. This set of indicators refers to the Capital adequacy (C), the Asset quality (A), the Management quality (M), the Earnings (E), the Liquidity (L) and the Sensitivity to market risk (S) of the institution. This representation helps the supervisory authority to identify institutions that are in need of attention. Our risk of distress model relies on proxies for CAMELS dimensions with a wide coverage. Among the possible balance sheet indicators we prefer to focus on those that are well-represented over time and across different types of institutions. In particular, capital adequacy is measured by the equity to assets ratio (Capital) and by the ratio of the sum of equity plus subordinated borrowings over total assets (Capital Funding Ratio). Capital adequacy represents the level of bank capitalization and higher values stand for better solvency conditions, thus lower values are expected to increase the probability of bank distress. The asset quality is assessed through the return on assets (ROA); in principle, better returns are negatively related to distressed conditions. We measure the management quality by means of the return on equity (ROE) and the ratio of operating expenses over operating income (Cost to Income Ratio). The relationships between management quality and the probability of distress is expected to be negative, as better management practices should foster economic performances and institution resilience. Net Interest Margin is utilised as a proxy for earnings and the expected sign of the relationship with distress is negative. In addition, the earnings dimension is approximated by the ratio of Interest Expenses to Total Liabilities and, in this case, the expected relationship is positive. Liquidity is measured by the ratio of liquid assets over customer and short-term funding (Liquid Assets to Short-Term Funding) and by the ratio

<sup>12</sup>For a study focused on US only, see the ranking-based methods of financial distress proposed by DeYoung and Torna (2013) aimed at revealing the identity of those US troubled institutions reported by FDIC.

of deposits and short-term funding over total funding (Deposits to Total Funding). Usually, institutions with better liquidity conditions are more likely to meet their financial obligations and thus are perceived as less risky (Tirole, 2011). Finally, the sensitivity to market risk is measured by the share of securities to total assets (Total Securities to Total Assets). The relationship with the probability of distress is ambiguous since securities are a volatile source of income but at the same time these assets can be more liquid than, for example, loans. This feature is particularly relevant for risk assessment during the recent crisis since the effects of fire sales, which represented a channel through which financial distress spread throughout the system, made some institutions more vulnerable. All indicators are computed with data retrieved from Bankscope.

In addition to the above individual banks characteristics, we consider a set of country-specific additional controls for financial sector and macroeconomic conditions (see e.g. Betz et al., 2014; Demirgüç-Kunt and Detragiache, 2005; Vazquez and Federico, 2015). Since our data set is characterized by a broad classification of financial institutions, we decide to represent the financial sector using a range of variables from banking indicators to market measures. This group includes domestic banking credit to non-financial sector (Credit to Non-Financial Sector), banking sector non-performing loans (Bank NPLs to Gross Loans), central government stock of debt (Central Govt. Deb), long-term government bond yield (Govt. Long-Term Yield), financial sector market returns (Sector Index), stock market returns (Market Index) and the amount of traded stocks (Stock Traded). Moreover, we take into account macroeconomic conditions by controlling for GDP per capita growth, Inflation, House Price, Unemployment rate and Foreign Direct Investments (FDI-Inflows and FDI-outflows). Data are retrieved from BIS, Datastream, OECD and World Bank.

Finally, to measure the impact of business model dynamics, such as the switch from a business model to another one, we add a categorical variable which represents the number of times (one or two) the institution changes peer group in the period 2005-07 (we label the variable *Switch Group*). We use this categorical variable to test whether a very volatile (stable) peer group membership increases (decreases) the probability of distress.

## 5 Methodology

As discussed in the Introduction, banking business models are usually identified indirectly by using the Ward and Joe (1963) clustering algorithm, or, alternatively, by means of direct classifications such as those provided by Bankscope. Literature based on the Ward approach exploits balance sheet variables (Ayadi et al., 2011; Roengpitya et al., 2014) to detect groups of homogeneous institutions. These studies test several combinations of assets and liabilities items to characterize banking activities and use euclidean distances to measure how similar financial institutions are with respect to these dimensions. Ward algorithm is usually combined with the Pseudo-F Index (Caliński and Harabasz, 1974) as a stopping rule to provide the number of clusters. The Pseudo-F Index is the ratio of between-cluster variance to within-cluster variance and is used as a metric to assess the quality of the clustering results and to discriminate among different specifications of the algorithm setup. In particular, the best configuration, and therefore the resulting number of clusters, is the configuration associated with the greatest value of the Pseudo-F Index. Typically, the selection of balance sheet items aims to preserve a balanced representation of both the assets and the liabilities sides opting for those variables with higher coverage among financial institutions. In our study we use a wider list of variables than previous studies and for a larger set of institutions globally distributed. For these reasons, we prefer to adopt a clustering method that, although in line with the Ward algorithm, is designed to address in a more appropriate and elegant way the multidimensionality issues, particularly severe in a large and sparse sample like ours (see Table 1). The following subsection presents our measure of similarity among financial institutions' balance sheet variables, i.e. the cosine similarities, and the hierarchical clustering algorithm we borrowed from complex system literature to identify peer groups, i.e. the *Louvain* community detection method (Blondel et al., 2008). Finally, the last Subsection discusses the Penalized-likelihood logistic model utilized to assess the risk of distress.

## 5.1 Peer Group Identification

Our classification method relies on the similarities among institutions' financial statement attributes, collected in a vector per bank and per year from 2005 to 2014. To determine these similarities we compute the cosine of the angle between each pair of vectors in the inner product space and we divide it by the vectors' L2 norms to make it bounded between  $-1$  and  $+1$ . The cosine similarity is a standard measure used in information retrieval (Dongen and Enright, 2012) which is typically applied for sparse and multidimensional data (Tan et al., 2006). Given two vectors  $x$  and  $y$ , their cosine similarity<sup>13</sup> is computed as follows:

$$\text{CosSim}(x, y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}} = \frac{\langle x, y \rangle}{\|x\| \|y\|} \quad (1)$$

Once we have measured the pair-wise similarities among institutions we apply a hierarchical clustering algorithm which identifies groups by maximizing the modularity quantity. This approach is common in complex system literature where the system resembles a graph or network of nodes (i.e. the financial institutions in our case) connected by means of edges (which link pairs of nodes/institutions and are weighted according to the similarities among them). Modularity measures the strength of division of a system into clusters or communities, where groups of densely interconnected nodes are only sparsely connected with the rest of the system (Newman and Girvan, 2004). The modularity is computed as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2)$$

where  $A_{ij}$  indicates the weight of the edge between nodes  $i$  and  $j$  (i.e. the similarity among this pair of institutions),  $k_i = \sum_j A_{ij}$  represents the sum of the weights of the edges attached to node  $i$  (basically it measures the similarity of institution  $i$  to the rest of the system),  $c_i$  is the cluster to which node  $i$  belongs (i.e. the peer group/business model),  $\delta(u, v)$  is equal to 1 if  $u = v$  and 0 otherwise, and  $m = \frac{1}{2} \sum_{i,j} A_{ij}$ . Among the approaches proposed in literature to optimise this quantity (hence, to provide a better partition of the system in clusters), we apply the *Louvain* method that has received an increasing interest in complex systems literature<sup>14</sup>. This algorithm is structured in two phases. Firstly, each institution is assigned to a single cluster, so there are as many clusters as there are institutions. Hence, for each institution the algorithm considers its neighbourhood and evaluates the gain of modularity which can be obtained by joining a different cluster. The combination that gives the maximum gain (if positive) is therefore performed. The process is repeated for all institutions until no further improvements are achieved. In particular, the gain in modularity (i.e.  $\Delta Q$ ) by moving an isolated institution  $i$  into a cluster  $c$  can be measured as follows:

$$\Delta Q = \left[ \frac{\sum_{in} + k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] \quad (3)$$

where  $\sum_{in}$  is the sum of the weights of the edges (i.e the similarities) within the cluster  $c$  and  $\sum_{tot}$  is the sum of the weights of the edges attaching to nodes in cluster  $c$ , while  $k_i$  is the

<sup>13</sup>Each component of the vector can be weighted according to the importance of that variable to the assessment of similarities among pairs of institutions. However, we adopt a neutral approach and we treat all information in the vector with the same importance to avoid ex-ante manipulation for the results.

<sup>14</sup>This part is indebted to the original paper by Blondel et al. (2008); we rely intentionally on their formulation for presenting the main characteristics of the algorithm. For a deep review of community detection methodologies, see e.g. Fortunato (2010).



sum of the weights of the edges received by node  $i$  and  $k_{i,in}$  is the sum of the weights of the edges from  $i$  to nodes belonged to cluster  $c$ , and  $m$  is the total sum of the weights of all the edges in the system. In the second phase the algorithm builds a hierarchical partition whose nodes are the clusters identified in the first step. Thus, the algorithm tries to re-apply the first phase to the resulting weighted system of *meta-nodes* and iterates this procedure until a maximum of modularity is achieved.

Actually, cosine similarities might assume positive or negative values (this is due to the fact that balance sheet items may be either positive or negative). The *Louvain* community detection algorithm requires that edges with higher values are assigned to stronger similarities. Therefore, we first get a distance metrics applying the metric preserving transformation  $\theta_{i,j} = \sqrt{0.5(1 - CS_{i,j})}$ , where  $\theta_{i,j} \in [0, 1]$  and  $CS_{i,j}$  is the cosine similarity between  $i$  and  $j$ , which ensures that similarities range between 0 and 1 (Dongen and Enright, 2012). Then, we define the value of pair-wise similarity between  $(i,j)$  as  $1 - \theta_{i,j}$ , so that pairs of institutions which are very similar receive higher weights. This approach is in line with Puliga et al. (2016) where a similar clustering strategy is applied to classify a smaller set of banks than ours and to detect the main economic features characterizing the emerging communities.

A few technical issues should be taken into account. First, for each year we prune the system by removing the edges below the 0.025 and above the 99.975 percentiles of the cosine similarity distribution. Second, since the system is very dense by construction, we remove redundant edges avoiding its fragmentation, i.e. keeping the system connected. We test several specifications by filtering edges below certain thresholds<sup>15</sup>. We recall that edges with higher values stand for higher similarity between pairs of institutions and the goal of the algorithm is to find clusters of similar institutions. We rely on the idea of finding such dense system in an Erdos-Renyi random graph, that is we maximize  $H = \sum_c M_c D(p_c||p)$ , where  $M_c$  stands for the number of possible edges in the community  $c$  (i.e.  $n_c(n_c - 1)/2$ ),  $p_c$  is the density of the community  $c$ ,  $p$  is the general density of the graph and  $D(x||y)$  is the binary Kullback-Leibler divergence (see Traag et al., 2013). Finally, among values of  $H$  which are candidates for being the maximum, we usually prefer those that present higher values of modularity unless it implies a tight pruning of the edges.

In addition, we provide a further analysis of the emerging clusters by testing the distributions of their constituents. Appendix A shows results from the non-parametric equality of medians tests (i.e. Kruskal-Wallis test) that is used to verify whether clusters originate from the same distribution. We consider a wide set of variables and test non-parametrically whether clusters differentiate from each other for each year in the interval 2005-14. Results indicate the presence of differences in medians which we have further analysed by means of post-hoc multiple pairwise comparisons (i.e. Dunn tests). Our choice of the algorithm to detect peer groups reflects the aim to rely on a clustering approach that is in line with previous and established literature on business model identification. Both the *Louvain* and the Ward methods are hierarchical clustering algorithms and the quantities they maximize to find clusters are somehow similar (modularity vs. between/within variances). Moreover, the non-parametric Kruskal-Wallis tests to detect differences among clusters and the multiple pairwise post-hoc comparisons, which we use to further verify that groups are distinct, resemble the Pseudo-F Index framework used in Ward to identify the best configuration of clusters.

Table 13 compares the *Louvain* results with those obtained by applying both the Ward method and the direct classifications provided by Bankscope. The quality of these three approaches is assessed using common measures borrowed from clustering validation techniques (see Halkidi et al., 2001 and Han et al., 2011), i.e. the average clusters' silhouettes, the Pearson Gamma coefficient and the ratio of the average within and the average between distances among institutions. Missing values represent a key issue in clustering algorithms whenever authors aim to accommodate the complete data set. The absence of certain balance sheet items is itself a sign of a business model feature and some authors like Ayadi and De Groen (2015) replace them with zeros. To compute the Ward algorithm we consider the entire set of variables listed in Table 1 and we fill missing values (NAs) using four criteria to guarantee that the complete data set is utilized. First, we fill NAs with zeros; second, we replace NAs by mean values of the corresponding variable in the sample; third, similarly to the previous case, we use medians instead of mean values; fourth, we use multiple imputations and we combine the Expected Maximization (EM) algorithm (Dempster et al., 1977)

<sup>15</sup>In particular, we filter the system according to thresholds from 0.7 to 0.5 using a decreasing step equals to 0.025 and, for each year, we select the threshold which maximizes the significance of the configuration.

used to find the mode of the posterior distribution with a bootstrap approach to take draws from this posterior (Honaker and King, 2010). In the latter case, we run 10 simulations for each year and we present average values for the selected clustering validation measures. However, thanks to its elegance in dealing with sparse data sets, the *Louvain* method does not require assumptions on missing values as by construction the cosine similarity treats them as unknown and bases the similarity analysis on available data only. Finally, clustering validation measures are computed for the *Louvain* method and the direct classification (i.e. Bankscope institutions’ specializations) using the resulting clusters from the respective approaches and, to enhance comparability, we fill missing values using the same four criteria as seen above.

## 5.2 Empirical Approach

Banks distresses during the recent financial crisis are investigated by means of a logit model on the cross-sectional distribution of three types of variables prior to the outbreak of financial markets of 2007. Our empirical model is:

$$Pr(Y_i = 1|\mathbf{x}_i) = \Lambda(\mathbf{x}_i\beta) \quad (4)$$

where  $\Lambda$  is a logistic function and  $Y_i$  assumes value 1 if institution  $i$  has been under distressed conditions between 2008-10 and 0 otherwise (see Section 4.2). In particular, vector  $\mathbf{x}$  includes bank-specific measures, financial sector indicators and macro variables. The choice of relying on three types of variables is in line with other works which aim to disentangle the effects of banks features from the impacts of sectoral dynamics and macro conditions (see Section 4.3). To limit endogeneity issues, we exploit the pre-crisis averages (from 2005 to 2007) of the explanatory variables as in Vazquez and Federico (2015). Finally, due to the presence of few distress events we decide to apply a rare event logistic regression to take into account the possibility of a small amount of cases of the rarer outcome. Hence, to reduce the small-sample bias in maximum likelihood estimation, we apply the Firth’s Penalized-likelihood logistic regression, which is a convenient approach to obtain finite and consistent estimates of regression parameters when maximum likelihood procedure suffers from complete or quasi-complete separation (Firth, 1993).

Since we are interested in capturing differences in the probability of distress across different business models we present several specifications of the main model. In particular, we run the analysis on the entire sample and we compare these estimates with those computed within peer groups separately. Finally, a very similar framework is applied for the analysis of the switching dynamics of institutions across peer groups during the interval 2005-07.

## 6 Global Business Models

This Section discusses the outcomes of our classification approach by investigating the business models we find, their features and geographic compositions as well as their evolution over time.

### 6.1 Clustering validation

We present the results of a series of non-parametric tests and multiple pairwise post-hoc comparisons aimed at validating our classification approach. It also compares our clustering outcomes against both the Ward method and the direct classification to test for **H0**. We compared groups according to a set of balance sheet measures (see Appendix A) selected as business models features. We recall that the clustering algorithm that we use maximizes the homogeneity within clusters compared to the rest of the system. In doing that, those institutions that are more similar in terms of the multidimensional vector of balance sheet attributes are grouped together. As discussed in Section 5, this implies the maximization of the *modularity* of the resulting configuration. In this Section we test whether and how these groups differ. This is in line with literature on banks business models identification which relies on other procedures like the Ward clustering technique to statistically test for differences in the emerging clusters (see e.g. the Pseudo-F index proposed by Caliński and Harabasz, 1974). In our case, comparisons have a dual goal: they first aim to

support the ability of the *Louvain* algorithm to group similar institutions by statistically testing for median differences in groups' balance sheet measures, and second they help the interpretation of emerging groups in terms of business models features on a large set of measures.

Appendix F.1 provides detailed results of the non-parametric equality of medians test (i.e. the Kruskal-Wallis test), which test whether groups originate from the same distribution. Tests are performed yearly from 2005 to 2014, separately for each measure. Results point to the presence of differences in medians which we are further analysed by means of post-hoc multiple pairwise comparisons (i.e. the Dunn test). We use the Bonferroni correction to take into account the FWER<sup>16</sup>. We use a restrictive significance level of 1% to judge whether two groups differ or not. In order to remove some noise due to the presence of potential outliers we also repeated the same analysis by focusing only on the institutions belonging to the 2nd and 3rd quantiles of the distribution for any given group/measure. This check aims to study groups' differences at the *core* of the distribution where membership to that group is more likely to be stable over time. Results show that groups on average have statistically significant peculiar characteristics which are also quite persistent over time. This supports the use of *Louvain* algorithm as a technique to identify peer groups. Furthermore, we discuss some examples in Appendix F.2 with multiple comparisons tests under extreme cases. In fact, the visualization of box-plots of balance sheet measures discloses peculiar patterns for some groups that can affect the pairwise comparisons. In the examples we present an intuitive way to check these cases. Results even in these scenarios are quite promising.

Appendix F.3 provides an additional description of the emerging peer groups. In particular, for each year and each peer group we report descriptive statistics for the average number of outlier values and missing values related to the measures used to build the vector to compute the cosine similarities (see Section 4.1). This is a further investigation of the homogeneity within each community. Results suggest that on average the presence of outlier values is circumscribed to few variables, while the issue of missing values is not a peculiar feature of any group.

Finally, we test the quality of our clustering approach against the Ward and the direct classification using main measures of clustering validation (see Appendix B). Results confirm that the direct classification is a poor indicator of peer group assessment whereas the indirect approaches (the *Louvain* adopted in this study and the Ward) provide far better clustering results across different model configurations and validation measures. The methodological advantage of the *Louvain* in dealing with sparse and complex data samples<sup>17</sup> (see Section 5.1) makes it a more suitable classification approach than the Ward for the task ahead<sup>18</sup>.

## 6.2 Economic Features of Business Models

We present the balance sheet features characterizing each business model, where each peer group is analysed in detail with respect to its composition and evolution over time. An analysis of the geographic map of business models that focuses on the banking activity in the main countries is discussed in Appendix D. Table 4 present summary statistics for the main economic aggregates used to characterise and distinguish business models prior to the crisis<sup>19</sup>. For a detailed representation of those statistics under different time intervals, see Tables from 14 to 16 in Appendix C.1. The following set of dimensions belongs to the balance sheet measures usually applied in literature to identify banks' business models (see Ayadi et al., 2011; Beltratti and Stultz, 2012; Lozano-Vivas and Pasiouras, 2010; Mergaerts and Vander Vennet, 2016; Roengpitya et al., 2014, among others).

As we can see when cross checking variables statistics across different time periods, we first notice that groups' characteristics are very stable over time. This result, in combination with the inter temporal stability of institutions within the same cluster shown in Table 5 (and discussed in

<sup>16</sup>We also employ other approaches, not reported, like Sidak and Holm corrections obtaining similar results.

<sup>17</sup>Studies on banking business models using Ward tend to select a much smaller sample of institutions (e.g. Roengpitya et al., 2014), mainly from the same geographic location (Ayadi et al., 2011; Ayadi et al., 2012; Ayadi and De Groen, 2015) and/or on a limited set of characteristics to minimize the issue of dealing with missing values.

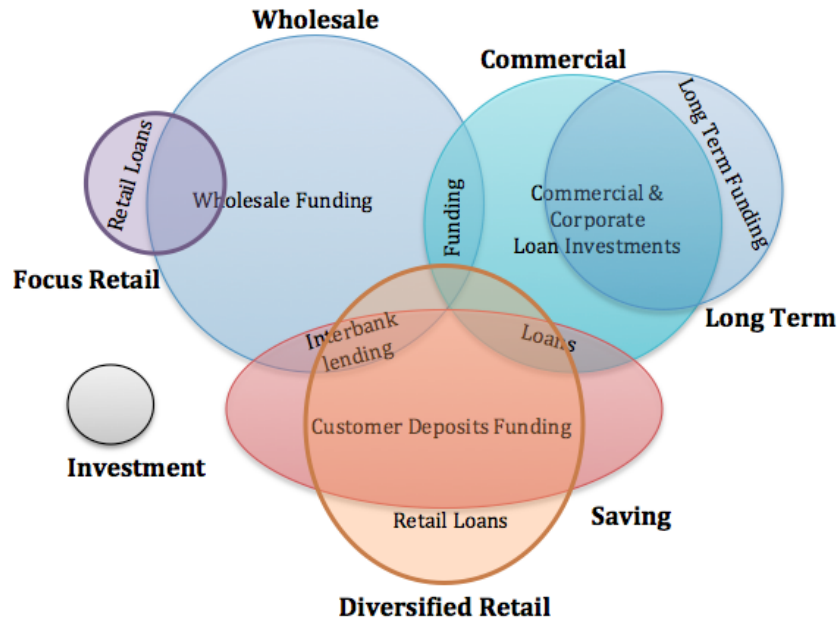
<sup>18</sup>As the core of this study is to discuss the economic implications of peer groups identification and their relations to the recent financial crisis, we leave the discussion of the comparisons among different clustering methods (already introduced in Appendix B), and in particular the parallel between our approach and Ward, to future studies.

<sup>19</sup>To provide a representation of the main features for each peer group, we prefer to rely on aggregate variables due to the presence of missing values among the measures used to compute the cosine similarities (for details see Appendix A).

detail in Section 6.4), confirms our findings of stable peer groups, thus supporting the interpretation of their features which are presented.

Our classification approach reveals the presence of seven peer groups representing the three business model categories also found in literature, such as the wholesale-oriented, the deposit-oriented<sup>20</sup> and the investment-oriented. Models names have been chosen on the basis of the particular characteristic that better discriminates each model against the others either from their funding or assets sides (Ayadi et al., 2012; Roengpitya et al., 2014). Detailed individual structures of each business model and comparisons with direct classifications are extensively discussed in Appendix C.

**Figure 1: Peer Groups.** Visualization of the peer groups. Intersection sets represents the common characteristics of the business models while the distinctive business model characteristics are listed in the centre of each group set. Sizes of the groups represents the popularity of business models measured by the total number of banks.



Three main business models are consistent throughout the entire reference period 2005-14 and accounting for the largest number of institutions (transitions of business models overtime are discussed in Section 6.4). The first of these models is characterised by the largest wholesale funding, accounting for about 32% of total assets on average, and well diversified loans investments with decent exposures to interbank activities. Many Russian commercial banks adopt this business model along with US bank holdings<sup>21</sup>. We name this model *Wholesale*<sup>22</sup> (see Appendix C.2 for details). The above figures are consistent with the “Wholesale-funded”<sup>23</sup> model discovered in Roengpitya et al. (2014) on a much smaller global sample. We also note substantial similarities with the “Wholesale”<sup>24</sup> business model presented in Ayadi et al. (2012) on a set of large European banks. Due to the important size of nondeposit funding as well as interbank activities, the *Wholesale* model presented here reflects interesting non-traditional banking activities that could jeopardize stability

<sup>20</sup>Many authors refer to this business model as retail-oriented since they focus on the funding side. However, to avoid confusion with our retail-based models that are named on the basis of the assets side (Retail Loans), we prefer to use this terminology as customer deposits are what characterized the retail-funded institutions.

<sup>21</sup>We include direct specializations provided by Bankscope to facilitate a parallel between our clustering approach and common classifications. Appendix C aims to deepen these aspects.

<sup>22</sup>In the paper we apply the convention to use *italics* for peer groups/business models identified by our clustering approach to differentiate these groups from the ones classified by the other methods.

<sup>23</sup>This model is characterised by a 65.2% of gross loans and 36.7% of wholesale debt, along with a 63.1% stable funding and 35.6% deposits, which are in the same range as our *Wholesale* model shown in Table 14. Their interbank composition is less prominent than ours, but still consistent across other models as we will show in the discussion of this Section.

<sup>24</sup>This model is characterised by interbank borrowings and lending (23.2% and 16.6% respectively) very similar to ours. Their Wholesale model tends to be more exposed to noninterest income investments and nondeposit funding than ours.

in distressed scenarios (see e.g. Lozano-Vivas and Pasiouras, 2010). This is the most popular group both in terms of number of institutions (over 3,000 on average) and total assets of the group (above \$60tn) with individual institution size above average, low leverage and mid net liquidity (decent level of maturity transformation).

The second model is mainly represented by US national commercial banks along with a large composition of European cooperative and saving institutions. Institutions are here characterised by a decent amount of wholesale funding and the largest exposure to commercial loans investments, which differentiate them from the well diversified *Wholesale* institutions. Due to the nature of their loan investments, we name this model *Commercial* (see Appendix C.3 for details). Institutions adopting this model tend to be smaller than the *Wholesale* ones with less than \$20tn of total assets (average on 2005-14) and individual size below average values. They also tend to be highly leveraged (with a thin equity capital buffer) with decent maturity transformation. This model resembles the risky Stakeholder banks reported in DeYoung and Torna (2013). It is interesting to note that Roengpitya et al. (2014), in their analysis of large 222 global banks, observe a wholesale-funded model, consistent with both our *Wholesale* and *Commercial* models, operated exclusively by non-US banks, mainly European. In our study, many US institutions adopting wholesale-oriented business models are bank holdings and national commercial banks, that were not part of their sample. However, the adoption of the *Commercial* model appears to be a distinctive characteristic of European banking systems as shown in Appendix D, which is in line with the findings in Roengpitya et al. (2014). The extension of the data set and the use of our indirect approach can, therefore, clarify banking activities across a wider range of institutional specializations for a more accurate analysis of business models both within and across countries.

The last core model is based on traditional customer deposits funding (the largest across models) invested in loans, mainly commercial, and some interbank lending exposures. The latter, in combination with low interbank borrowings, makes those institutions the only interbank net lenders among the three core models as shown in Table 4. Many US, Japanese and Indian commercial banks adopted this model. We name it *Saving* (see Appendix C.4 for details). We note that Ayadi et al. (2012) find a deposit-oriented model to be net-borrower in the interbank market as opposed to our *Saving* model. Their result is actually in line with another deposit-oriented model we find, i.e. the *Diversified Retail* group, that is introduced below and further discussed in Appendix C.5. In fact, the latter shows a better European representation than the *Saving* model and could justify the similarity with Ayadi et al. (2012) as their analysis is restricted to European institutions only. The *Saving* group provides on paper the most stable funding business model due to the deposit-driven liabilities. Roengpitya et al. (2014) find that their retail-funded model, closely related to our *Saving* model<sup>25</sup>, presents less volatile earnings compared with the other models, although it is not as cost-efficient as the wholesale-funded one. The *Saving* model is also the second largest group in terms of total assets (above \$20tn on average on the period 2005-14) with individual institutions size above average. High leverage along with dominant commercial loan investments makes this model very similar to the *Commercial* one, although featuring a much more stable funding due to the large proportion of customer deposits to total funding. This is also the model that provides the most dominant maturity transformation of all (the lowest net liquidity value), confirming the traditional banking service provided. This feature confirms the empirical evidence reported in Paligorova and Santos (2016) on the maturity transformation attitude of wholesale-oriented compared with deposit oriented banks.

The institutions within each of these three core models well represent the two main accounting standards, GAAP and IFRS, suggesting that our classification approach does not show clear biases of accounting manipulation<sup>26</sup>.

The 2007-08 financial crisis produced what can be interpreted as a huge earthquake in the balance sheets of institutions all around the world, which therefore influenced the adoption of business models as also reported in Roengpitya et al. (2014). Figure 1 visualize the similarities and differences between peer groups. On top of our three core models introduced above, our classification

<sup>25</sup>The retail-funded model found by Roengpitya et al. (2014) is characterized by the largest deposit driven model with 66.7% of total assets, very high stable funding and quite diversified assets side in line with our *Saving* model.

<sup>26</sup>Our database includes 65.4% of institution under GAAP standards and 23.4% IFRS in 2005 (the remaining are mainly US institutions under regulatory accounting standards), marginally moving towards a 60% and 32% composition respectively in 2014. Both accounting standards are well represented in all three core models, with a marginal preference of GAAP in the *Wholesale* and *Saving* business models and IFRS in the *Commercial* one.



approach captures two relevant models that were popular only in the first two years of our sample and then disappeared at the onset of the crisis. These wholesale-oriented business models were adopted by a smaller number of institutions as reported in Table 4. The first, which we named *Long Term* model, was characterized by dominant long-term funding and commercial investments. Many European institutions, in particular Italian and Spanish banks, adopted this model (see Appendix C.6). Other peculiarities of this model is the net interbank lending exposure that is shared with the *Saving* model. They account for about 1,500 institutions in the two years of the existence of the group with total assets size greater than the *Commercial* model. Due to the close similarity to the *Commercial* model with regards to their assets side, we anticipate that the majority of these institutions migrated to the *Commercial* business model in 2007 with few exceptions to the *Saving* model as discussed in Section 6.4. The second model, which we labelled *Focus Retail*, shows a diversified funding combined with the largest exposure to retail loans investments. Institutions belonging to this group were mainly Swiss institutions and US bank holdings. Those banks are also highly leveraged compared with the other groups. We also report quite a big variation in the distribution of total assets within the group (see Appendix C.7 for details). At the onset of the crisis most of them migrated to the *Wholesale* peer group.

**Table 4: Peer Groups Economic Features - Pre Crisis Summary Features.** We report average values for aggregated balance sheet variables standardized by total assets for groups Wholesale, Commercial, Saving, Long Term and Focus Retail. Column *Average (I-III)* refers to the average values within the sample composed by groups Wholesale, Commercial and Saving. Column *Average (I-V)* refers to the average values within the sample composed by all five groups. For variables definitions see Appendix A. For groups Long Term and Focus Retail, estimates refer to the interval 2005-06 (since they disappear in 2007), while for Wholesale, Commercial and Saving we consider the interval 2005-07. Last row provides summary statistics for Total Assets (in USD Billion). Source: Bankscope, Authors' own elaborations.

		Wholesale (I)	Commercial (II)	Saving (III)	Long Term (IV)	Focus Retail (V)	Average (I-III)	Average (I-V)
	# observations	3573	3007	1989	1475	535	2856	2116
Retail Loans	<i>Mean</i>	0.17	0.01	0.02	0.01	<b>0.73</b>	0.08	0.10
Corporate and Other Loans	<i>Mean</i>	0.36	<b>0.61</b>	0.54	0.59	0.07	0.49	0.48
Retail and Corporate Loans	<i>Mean</i>	0.52	0.62	0.56	0.60	0.80	0.57	0.58
Total Loans	<i>Mean</i>	0.72	0.72	0.71	0.75	0.87	0.72	0.73
Interbank Lending	<i>Mean</i>	0.19	0.10	0.15	0.15	0.07	0.15	0.15
Investments	<i>Mean</i>	0.17	0.22	0.21	0.17	0.09	0.20	0.19
Customer Deposits	<i>Mean</i>	0.40	0.48	<b>0.79</b>	0.57	0.49	0.52	0.52
Interbank Borrowing	<i>Mean</i>	0.25	0.16	0.06	0.08	0.10	0.17	0.15
Long-Term Funding	<i>Mean</i>	0.10	0.06	0.03	<b>0.18</b>	0.12	0.07	0.09
Long-Term Funding + Equity	<i>Mean</i>	0.25	0.13	0.13	0.30	0.16	0.18	0.20
Wholesale Debt	<i>Mean</i>	<b>0.34</b>	0.22	0.09	0.26	0.22	0.24	0.24
Stable Funding	<i>Mean</i>	0.57	0.56	0.84	0.78	0.62	0.63	0.65
Net Liquidity	<i>Mean</i>	-0.59	-0.61	-0.81	-0.61	-0.57	-0.65	-0.64
Total Assets (USD Billion)	<i>Mean</i>	20.77	9.00	13.37	16.91	12.00	14.92	15.05

At the peak of the financial turmoil, we observe the emergence of a peculiar and large group characterised by a business model with dominant customer deposits funding, second largest after the *Saving* model, and dominant exposure to retail loans investments. We name this model *Diversified Retail* (see Appendix C.5). This group is composed by mainly German and Swiss saving and cooperative institutions which clusterised together after 2008 in this new business model that persisted thereafter. This model is the result of a substantial restructuring process during the outbreak of the financial markets, aiming at improving their stable funding, which indeed increased from about 53% (average of *Wholesale* and *Commercial* models) to 77%. This model is characterised by a hybrid liabilities structure that incorporates features of both *Saving* and *Wholesale* models. Statistical information and comparisons with the three core models are given in Table 16. Similar to European institutions moving from *Long Term* to *Saving* prior to the crisis, we



remark this interesting migration as evidence of institutions transforming their wholesale-oriented funding models into more traditional deposit-based activities as also reported in Roengpitya et al. (2014). This model becomes very popular with more than 2,000 institutions joining from the other wholesale-oriented groups, although their individual size is well below the average (due to the vast majority of small saving and cooperative banks populating the group). This model also shows high leverage and low net liquidity suggesting a high degree of maturity transformation in line with deposit-oriented models. Supported by the findings of DeYoung and Torna (2013), Demirgüç-Kunt and Huizinga (2010) and Köhler (2015), the impact of variations of non-interest income and non-deposit funding to the risk of distress is discussed in Section 7 by focusing on the changes of business models prior to the crisis between deposit and wholesale-oriented ones.

Finally, two small and residual models appear at the end of the sample period suggesting a slower re-organization of the financial system after the outbreak of financial markets of 2007-08 and consequent changes in the regulatory framework as well as banking practices. Although one of them shows a volatile composition of institutions which resembles a transient group (making the interpretation of the adopted business model quite difficult, see Appendix C.9), the second model points to specialized non-traditional activities with low customer deposits funding, high levels of interbank and wholesale borrowings and large exposure to both non-interest income investment and interbank lending (see Table 16). We named this model *Investment*, which appears only in 2012 and 2014. The huge size of the institutions adopting this model is a distinctive characteristic of the group, which is represented by the major international broker dealers, mainly from UK, US and Japan (see Appendix C.8). These institutions provide nontraditional banking services with decent leverage and very high net liquidity (low maturity transformation).

We conclude this Subsection with a note on the direct specialization of the institutions. Appendix C show how misleading the direct specialization of a bank is in representing its true business model (evidence confirmed by our clustering validation tests in Section 6.1). Although we observe within countries similarities between direct and indirect classification, our finding clearly outline the comparability discrepancies of banks direct classification in a global contest (e.g., EU saving and cooperative banks adopting same business models as US commercial and investment banks).

### 6.3 Size Effect

We now compare the distribution of total assets of distressed institutions with that of their peer group. Figure 2 displays this comparison by plotting the time series of the average size and volatility of institutions that faced distress in 2008-10 (black lines, blue dispersion) against those of the whole peer group (red lines, grey dispersion). Our findings suggest that all wholesale-oriented peer groups show an interesting size effect of distressed institutions being much bigger in terms of total assets than their peers.

Almost half of the distress events we collect globally come from institutions adopting the *Wholesale* model at some points in the period 2005-07 (precisely: 72 in 2005, 75 in 2006 and 92 in 2007; 64 adopting continuously the model during the interval 2005-07). These distressed institutions, although few in numbers compared to the total, account for a huge portion of the total assets of the peer group, such as \$10tn in 2005 with an average institution size of \$132.2bn, seven times larger than the average size of the group (if we restrict to institutions always belonging to the *Wholesale* model in the interval 2005-07 the average size of distressed institutions is \$135.5bn). Even restricting the comparison per country (to account for country effects) we still observe that being bigger relative to the average of peer group members, even from the same country, exacerbates vulnerability and eventually distress. Section 7 empirically tests the contribution of individual size to the likelihood of distress.

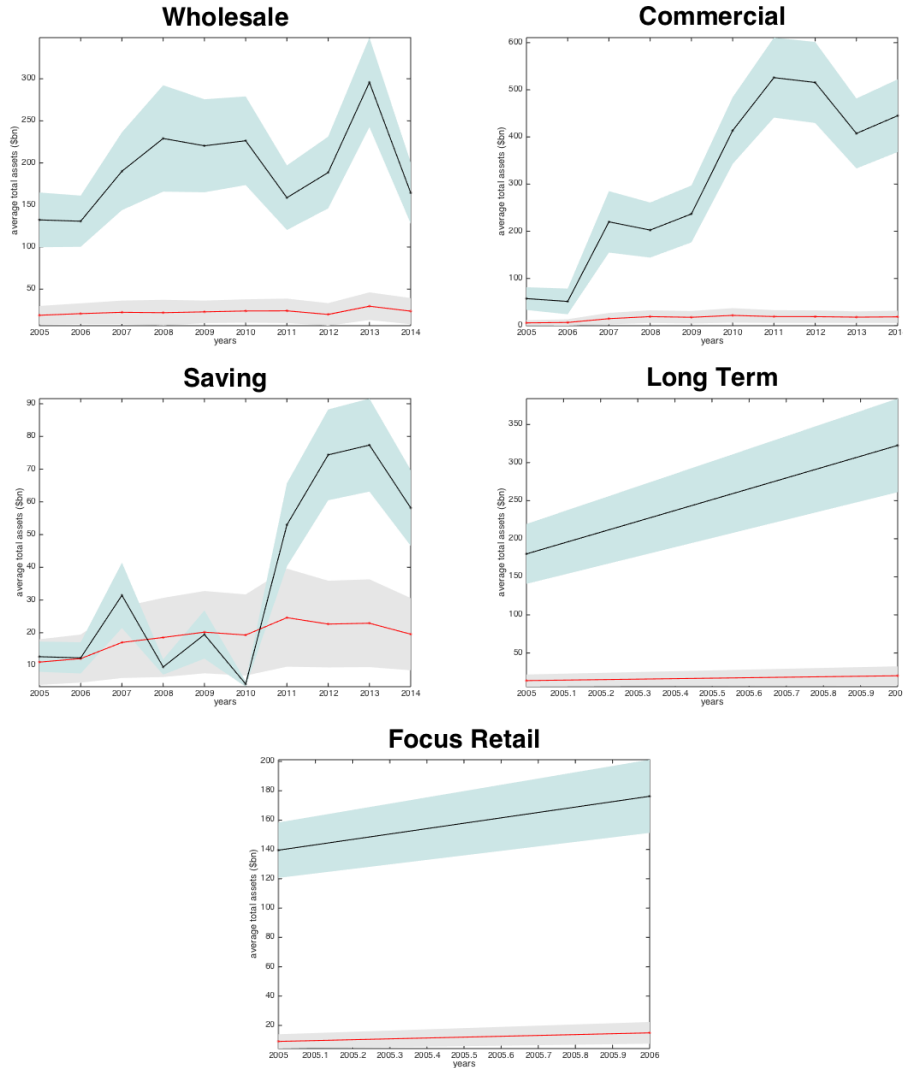
*Commercial* model exhibits a smaller number of institutions in distress during the crisis (33, 35 and 55 distressed institution which belong to this group in 2005, 2006 and 2007 respectively; 28 steadily part of the *Commercial* group during the interval 2005-07) than the *Wholesale* model. This result is also justified by the lower popularity of the *Commercial* versus the *Wholesale* model. However, same evidence of larger size on vulnerable banks compared to their peers is observed. For example, the 33 institutions in 2005 that will show distress during the crisis period are on average 10 times bigger than their peers, even by restricting the comparison within the same country of origin. They account for almost \$2tn of assets, almost 15% of the total of the group. This evidence

also supports the claim that relative size matters for vulnerability and probability of distress.

The *Saving* group differentiates itself from the other two core groups (and from any other wholesale-oriented as we can see in Figure 2) on the risk attitude of the institutions. There is in fact no evidence of higher vulnerability of those on the right side of the distribution of assets relative to their peers. The average size of distressed institutions during the period 2008-10 in this group is perfectly in line with their peers, with non consistent patterns to be spotted at the country level either. This would suggest that relative size on a model with very high stable funding does not matter for purposes of risk assessment. This effect is tested in Section 7.

The risk profile of distressed institutions in the *Long Term* group is quite peculiar: 27 institutions out of 1416 in 2005 were under distress during the financial crisis (the number increases to 30 out of 1534 in 2006). Those few represented a total assets of almost \$5tn in 2005 (\$9.7tn in 2006), almost a forth and a third of the total peer group's assets in 2005 and 2006, respectively. Those 27 institutions had an average size back in 2005 of \$180bn (\$322bn in 2006), more than 13 and 16 times the average size of their peer group members in 2005 and 2006, respectively. This places those vulnerable institutions at the far high end of the distribution of assets, providing evidence of the size effect found in other wholesale-oriented groups.

**Figure 2: Total Assets Distribution** Red curve represents the time series of average total assets (in US Billion) for all institutions in each peer group, while blue curve stands for distressed institutions only. Dispersion area refers to  $\pm 0.1\sigma$ . Membership is updated yearly. Source: Bankscope, Authors' own elaborations.



To conclude with the last wholesale-oriented model that do not exceed in large stable fundings, the *Focus Retail* model shows distress events mainly among few very large institutions (13 and 12 institutions that adopted this model in 2005 and 2006, respectively), accounting about one third of the total assets of the group. Average sizes of these vulnerable institutions (mainly from UK, the Netherlands and Switzerland) are \$139.5bn in 2005 and \$176.2bn in 2006, 15 times larger than the average size of their group members in 2005 and 12 times for 2006.

## 6.4 Transitions across Business Models

In this Section we focus on financial institutions' transitions across business models. This analysis allows us *i)* to study the stability of business models in the interval 2005-14 and *ii)* to advance some possible explanations of institutions' changes in their peer group membership (which is discussed in Section 7.3) that would provide supporting evidence for **H2** and **H2bis**. Knowing that institutions tend to run consistent business models over time, a very low transition probability would validate our identification of peer groups (see Appendix F for a detailed analysis of year-by-year representation of the business models).

Figure 3 exhibits that institutions tend to persist in the same peer group during the biennium 2005-06, while the disappearance of both *Long Term* and *Focus Retail* groups in 2007 determined a migration of these banks into the three core models. Consistently with their funding orientations, institutions belonging to *Long Term* and *Focus Retail* models in 2006 migrated mainly to the other two available wholesale-oriented models, i.e. *Wholesale* and *Commercial* groups. We note that almost all *Focus Retail* institutions (98.2%) moved to the *Wholesale* group, most probably for the assets side diversification that the *Wholesale* model offers to institutions that used to have 73% of their assets invested in retail loans (see Table 15). A different dynamics affects institutions in the *Long Term* group, which predominantly migrated to the *Commercial* model due to the similarity between their assets structures. Only 16.7% of *Long Term* institutions moved to the *Wholesale* model and, as expected, just few (3.8%) converted drastically to the deposit-oriented peer group.

**Figure 3: Business Models Transitions.** The plot shows the percentages of institutions belonging to a certain model group in one period and switching to another model in the next period. Plot on left refers to switches from 2005 to 2006, while plot on the right is for transitions from 2006 to 2007. Source: Authors' own elaborations.

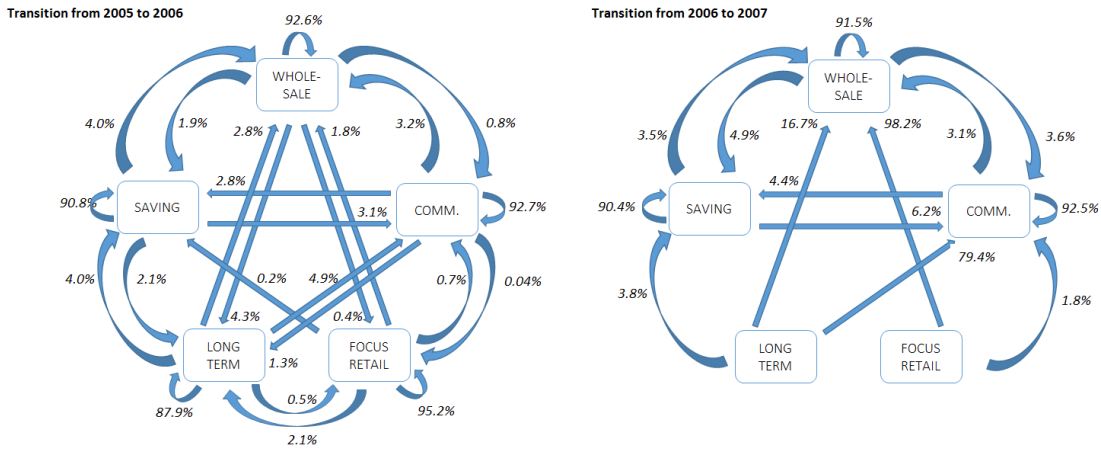


Table 5 provides the membership stability over time, that is the percentage of institutions that kept the same business model from the previous year. With an average of almost 90%, membership to business model seems to be quite stable over the period 2005-14, thus confirming that these models are basically composed by a constant set of institutions during the reference period. This result also validates the effectiveness of our peer group assessment on the inter-temporal dimension. We find that deposit-oriented models tend to be more stable, with in particular the *Diversified Retail* (mainly Swiss and German banks) one of the most cohesive model due to probably its geo-

graphic homophily. We also notice a breakpoint in correspondence of the collapse of 2007. In this year the three core models, and especially the wholesale-oriented groups, were *contaminated* by the inflows of institutions from the other two groups. Although one might argue that the resulting three groups in 2007 and hereinafter are no longer the same as the ones emerged in 2005-06, we still observe in Appendix F.1 a reasonable continuity in the distributions of balance sheet features around the crisis of 2007. Wholesale-oriented models are those more affected by the inclusion of institutions belonging to different groups in the biennium prior to the crisis, however percentages shown in Table 5 indicate that still in 2007 these models maintain a high proportion of members which belonged in 2006 to the same peer groups. Finally, it is worth underlining that stability of business models over time in terms of balance sheets characteristics is obviously a high desirable requirement for a reasonable clustering algorithm, although a certain degree of variability might be due to the normal updating process of banking activities. Since this sample period includes one of the most significant event in the sustainability of the financial markets, it seems realistic that institutions reacted differently from the past against the deteriorated market conditions and that business models have been greatly influenced by the wave of financial turmoil. This, in turn, poses several issues in the recognition of consistent business models and in the assessment of the coherence of these groups over time. In particular, before the breakdown of financial market in 2007 institutions experienced a high level of deregulation and financial innovation, while after the onset of the crisis the establishment of a new regulatory framework (e.g. the Basel III regulations and the Dodd-Frank Act) as well as macro and micro prudential decisions pointed to a more robust and regulated financial system. Our results indicate the presence of three main business models which persist during the entire interval 2005-14 and that present quite stable balance sheet figures. Remarkably, our peer group assessment also identifies a convergence to three main business models at the outbreak of financial market and the departure from 2008 of a homogeneous group (*Diversified Retail*) either in term of balance sheets features and geographic coverage.

**Table 5: Model Membership Stability over Time.** In this table we provide the percentage of institutions that confirm the same business models in time  $t$  with respect to their model in  $t-1$ . Source: Authors' own elaborations.

Group	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
Wholesale	93.08%	76.14%	92.72%	92.69%	86.99%	85.88%	88.33%	86.16%	76.34%	86.48%
Commercial	93.98%	63.47%	85.86%	81.19%	90.57%	89.83%	93.85%	91.00%	81.22%	85.66%
Saving	90.25%	84.57%	92.81%	75.46%	93.41%	92.91%	85.75%	94.22%	93.75%	89.24%
Diversified Retail				81.99%	96.53%	95.74%	94.64%	95.77%	96.88%	93.59%
Long Term	85.19%									
Focus retail	95.87%									

## 7 Risk Assessment

This Section analyses institutions' resilience to the risk of distress arising from the outbreak of financial markets in 2008. We apply a Penalized-likelihood logistic regression to study the distress events presented in Table 2 (see Section 4.2). Following Betz et al. (2014), we exploit three sets of information as regressors, i.e. individual characteristics captured by CAMELS proxies as well as macroeconomic and sectoral control variables which are introduced in Section 4.3 (see Table 3). Summary statistics of regressors' distributions are provided in Table 6. Due to the nature of these regressors, we might suspect the presence of potential multicollinearity issues. Table 7 shows that correlations over the period 2005-07 among regressors do not exhibit relevant relationships, except for few cases, and estimates are usually significant. In addition, we present different specifications of the baseline model for risk assessment, where we basically select subsets of regressors to overcome potential issues related also to multicollinearity.

**Table 6: Regressors Summary Statistics.** Values are computed as averages over the interval 2005-07. For variables definitions see Section 4.3. Source: our elaborations on data from Bankscope, BIS, Datastream, OECD and World Bank.

	obs	min	1st qu.	median	mean	3rd qu.	max
<i>Capital</i>	7906	-0.22	0.06	0.08	0.11	0.13	0.99
<i>Capital Funding Ratio</i>	7906	-1.19	0.06	0.09	0.12	0.13	0.99
<i>Roa</i>	7854	-0.24	0.00	0.01	0.01	0.01	0.36
<i>Cost to Income Ratio</i>	8376	0.01	0.56	0.66	0.65	0.74	5.79
<i>Roe</i>	7848	-1.46	0.04	0.08	0.09	0.14	1.60
<i>Net Interest Margin</i>	7825	-0.48	0.02	0.03	0.04	0.04	2.66
<i>Interest Expenses to Total Liabilities</i>	7769	-0.01	0.02	0.02	0.03	0.03	14.29
<i>Liquid Assets to Short-Term Funding</i>	7784	0.00	0.11	0.21	0.34	0.40	7.73
<i>Deposits to Total Funding</i>	7857	0.00	0.84	0.96	0.88	1.00	1.16
<i>Total Securities to Total Assets</i>	8427	0.00	0.07	0.17	0.19	0.27	1.00
<i>GDP per capita</i>	8509	-0.08	0.02	0.03	0.03	0.03	0.27
<i>Inflation</i>	8274	0.00	0.02	0.02	0.03	0.03	0.30
<i>House Price</i>	7322	-0.10	-0.01	0.01	0.04	0.04	0.31
<i>Unemployment</i>	8464	0.01	0.05	0.07	0.07	0.10	0.36
<i>FDI-Inflows</i>	8476	-0.06	0.02	0.02	0.04	0.04	0.57
<i>FDI-Outflows</i>	8239	-0.02	0.02	0.03	0.08	0.04	3.81
<i>Central Govt. Debt</i>	7406	0.04	0.42	0.44	0.58	0.62	1.81
<i>Govt. Long-Term Yield</i>	7495	0.02	0.04	0.04	0.04	0.05	0.15
<i>Bank NPLs to Gross Loans</i>	8128	0.00	0.01	0.03	0.03	0.03	0.55
<i>Credit to Non-Financial Sector</i>	7457	0.10	0.54	0.83	0.79	0.92	1.63
<i>Market Index</i>	7769	-0.05	0.07	0.17	0.21	0.28	0.74
<i>Sector Index</i>	7466	0.06	0.20	0.20	0.31	0.26	1.32
<i>Stock Traded</i>	7020	0.00	0.73	0.77	1.24	2.24	5.26

**Table 7: Correlation Matrix.** The upper triangular matrix is computed using observations corresponding to averages between 2005-07 (complete cases). In the lower triangular matrix we allow for the presence of one missing value in the interval 2005-07 when we compute the average values. \* refers to significant correlations at 1% level. Source: our elaborations on data from Bankscope, BIS, Datastream, OECD and World Bank.

	Capital	Cap. Fund. Ratio	Rea	Cost to Inc. Ratio	RoE	Net Int. Margin	Int. Exp. to Tot Liab.	Int. Assets to ST Fund.	Dep. to Tot Fund.	Tot Sec. to Tot Assets	GDP per Capita	Inflation	House Price	Unempl.	FDI-In.	FDI-Out.	Cont. Govt. Debt	Govt. LT Yield	Bank NPLs to Gross Loans	Credit to Non-Fin. Sector	Mlt Index	Sector Index	Stock Traded	Total Assets
Capital	0.9513*	0.4743*		-0.0482*	0.0444*	0.2249*	0.0803*	0.4300*	-0.0670*	0.0076	0.2395*	0.3188*	0.3455*	0.0210	0.0492*	-0.0315*	-0.1314*	0.2813*	0.0392*	-0.2453*	0.3024*	0.3127*	0.0171	-0.0813*
Capital Funding Ratio	0.9513*		0.4387*	-0.0538*	0.0369*	0.2224*	0.0615*	0.4200*	-0.1077*	-0.0089	0.2240*	0.3179*	0.3429*	0.0232	-0.0502*	-0.0299*	-0.1761*	0.2369*	0.0377*	-0.2486*	0.2650*	0.3048*	0.0255	-0.0728*
Rea	0.4745*	0.4380*		-0.3432*	0.3950*	0.2290*	0.0106	0.2124*	-0.0622*	0.0370*	0.1486*	0.2559*	0.2465*	-0.0275	0.0806*	0.0043	-0.1637*	0.2287*	0.0099	-0.1778*	0.1575*	0.2098*	0.0183	-0.0261
Cost to Income Ratio	-0.0323*	-0.3564*		-0.4310*	-0.0187	-0.0543*	-0.0543*	0.0180	0.1203*	0.0194	-0.0310*	-0.0524*	-0.0710*	0.0574*	-0.1088*	-0.0634*	0.0510*	-0.0174*	-0.0129	-0.0052	-0.0470*	-0.0309	-0.0208	-0.0295
RoE	0.0446*	0.0373*	0.5715*		0.1397*	0.1397*	0.0073	0.0694*	-0.0692*	-0.0263	0.1449*	0.2293*	0.2230*	-0.0090	0.1138*	0.0494*	-0.1066*	0.2343*	-0.0128	-0.1559*	0.1321*	0.1598*	0.0130	0.0470*
Net Interest Margin	0.2254*	0.2227*	0.2293*	-0.0194	0.1395*		0.0282	0.0496*	-0.0365*	-0.1416*	0.2212*	0.3215*	0.2481*	0.0660*	-0.0326*	-0.0691*	-0.1415*	0.2919*	0.0740*	-0.2661*	0.2415*	0.2892*	-0.0686*	-0.0596*
Interest Expenses to Total Liabilities	0.0986*	0.0908*	0.0103	-0.0189	0.0066	0.0283		0.0512*	-0.0730*	-0.0016	0.0258	0.0547*	0.0270	0.0088	0.0052	0.0000	-0.0037*	0.0561*	0.0127	-0.0590*	0.0276	0.0246	0.0041	-0.0042
Liquid Assets to ST Funding	0.4510*	0.4401*	0.2241*	0.0190	0.0613*	0.0381*	0.0858*		-0.2668*	0.0268	0.1332*	0.1796*	0.2313*	-0.0031	0.1120*	0.1062*	-0.0642*	0.1053*	0.0150	-0.0809*	0.1529*	0.2397*	-0.0865*	0.0563*
Deposits to Total Funding	-0.0679*	-0.1087*	-0.0651*	0.1156*	-0.0334*	-0.0322*	-0.0175	-0.2866*		0.0494*	-0.0097	-0.1066*	-0.2345*	0.0726*	-0.0311*	0.0206	-0.0245	-0.1166*	-0.0722*	0.1653*	0.0044	-0.1568*	-0.0098	-0.1466*
Total Securities to Total Assets	0.0076	-0.0127	0.0331*	0.0094	-0.0396*	-0.1435*	-0.0017	0.0380*	0.0495*	0.0494*	-0.1877*	-0.1701*	-0.2563*	-0.0084	-0.0536*	-0.0112	0.1801*	-0.1258*	-0.0113	-0.0038	-0.1401*	-0.2313*	-0.0205	0.1448*
GDP per Capita	0.2199*	0.2240*	0.1482*	-0.0503*	0.1430*	0.2284*	0.0256	0.1385*	-0.0095	-0.1887*	0.7167*	0.7192*	0.7123*	0.2048*	0.2080*	0.0379*	-0.5484*	0.5556*	0.0896*	-0.3251*	0.7596*	0.8382*	-0.2047*	-0.0425*
Inflation	0.3189*	0.3180*	0.2555*	-0.0496*	0.2327*	0.3221*	0.0548*	0.1644*	-0.1060*	-0.1777*	0.7167*	0.7192*	0.8147*	0.1033*	0.0993	-0.0577*	-0.5239*	0.5550*	0.1870*	-0.6856*	0.6404*	0.8349*	-0.0980*	-0.0527*
House Price	0.3458*	0.3429*	0.2462*	-0.0728*	0.2150*	0.2481*	0.0271	0.2100*	-0.2311*	-0.2566*	0.7133*	0.8147*	0.8147*	0.1033*	0.0993	-0.0577*	-0.5239*	0.5550*	0.1870*	-0.6856*	0.6404*	0.8349*	-0.0980*	-0.0527*
Unemployment	0.0210	0.0232	-0.0278	0.0584*	-0.0075	0.0965*	0.0088	-0.0084	0.0728*	-0.0111	0.2948*	0.0990*	-0.0429*	-0.0738*	-0.0775*	-0.1053*	-0.2613*	0.1332*	0.3113*	-0.0429*	0.4127*	0.0811*	-0.4028*	-0.0468*
FDI-Inflows	0.0300*	0.0361*	0.0373*	-0.0655*	0.0580*	-0.0211	0.0045	0.0577*	-0.0444*	-0.0311*	0.2903*	0.0907	0.0967*	-0.0738*	0.2538*	0.2538*	-0.1975*	-0.0186	-0.0350*	0.2466*	0.0652*	0.0446*	-0.0320*	0.0749*
FDI-Outflows	-0.0325*	-0.0310*	0.0028	-0.0650*	0.0463*	-0.0705*	0.0000	0.0897*	0.0203	-0.0014	0.0361*	-0.0590*	-0.0099	-0.1040*	0.2520*	-0.1107*	-0.1790*	-0.1107*	-0.1028*	0.0539*	-0.0548*	-0.0302*	-0.1292*	0.0085
Central Govt. Debt	-0.1522*	-0.1790*	-0.1692*	0.0544*	-0.1803*	-0.1455*	-0.0401*	-0.0643*	-0.0233	0.1828*	-0.5362*	-0.5191*	-0.4134*	-0.2531*	-0.1992*	-0.1778*	-0.5727*	-0.5727*	0.2052*	0.1300*	-0.6113*	-0.0353*	-0.0693*	0.0275
Govt. Long-Term Yield	0.2813*	0.2986*	0.2779*	-0.0425*	0.2265*	0.2019*	0.0662*	0.0984*	-0.1159*	-0.1296*	0.5566*	0.8580*	0.6121*	0.1332*	-0.0186	-0.1107*	-0.5727*	0.0847*	0.847*	-0.7421*	0.5566*	0.6169*	0.0665*	-0.0280
Bank NPLs to Gross Loans	0.0366*	0.0401*	0.0149	-0.0111	-0.0102	0.0941*	0.0125	0.0081	-0.0656*	-0.0068	0.1055*	0.2038*	-0.0160	0.3166*	-0.0428*	-0.1055*	0.1967*	0.0884*	-0.1680*	-0.1651*	0.3250*	0.1080*	-0.4397*	-0.0411*
Credit to Non-Financial Sector	-0.2453*	-0.2480*	-0.1769*	-0.0119	-0.1517*	-0.2663*	-0.0592*	-0.0725*	0.1658*	-0.0017	-0.3251*	-0.0856*	-0.4333*	-0.0429*	0.2466*	0.0309*	0.1306*	-0.7421*	-0.1680*	-0.1651*	-0.2536*	-0.4472*	-0.0624*	0.0572*
Market Index	0.1866*	0.1919*	0.1379*	-0.0343*	0.1231*	0.2377*	0.0275	0.1289*	0.0034	-0.1422*	0.7426*	0.6251*	0.6124*	0.4093*	0.0096*	0.0647*	-0.5003*	0.3441*	0.3126*	-0.2569*	0.7698*	0.7698*	-0.5769*	-0.0639*
Sector Index	0.3127*	0.3048*	0.2098*	-0.0265	0.1479*	0.2804*	0.0247	0.2137*	-0.1565*	-0.2311*	0.8262*	0.5549*	0.9031*	0.0811*	0.0446*	-0.0510*	-0.4022*	0.6169*	0.1095*	-0.4472*	0.7698*	0.7698*	-0.5769*	-0.0639*
Stock Traded	0.0694	0.0179	0.0145	-0.0163	0.0043	-0.0770*	0.0033	-0.0769*	-0.0112	-0.0182	-0.2713*	-0.1218*	-0.0359*	-0.4795*	-0.0450*	-0.1229*	-0.0860*	0.0370*	-0.0621*	-0.0621*	-0.5797*	-0.6282*	-0.6289*	0.0280
Total Assets	-0.0813*	-0.0720*	-0.0264	-0.0284	0.0448*	-0.0590*	-0.0042	0.0179*	-0.1410*	0.1451*	-0.0425*	-0.0528*	-0.0123	-0.0468*	-0.0335*	0.0087	0.0279	-0.0280	0.0372*	-0.0635*	-0.0635*	-0.0517*	0.0299	



**Table 8: Distribution of Distress Events for Model Membership.** Column names refer to peer groups in year 2007; row names in bold stand for peer groups in year 2006; finally, row names in italics refer to peer groups in 2005. Each cell represents the number of distressed institutions referring to the corresponding combination of peer groups membership. Row *Total (2007)* shows the number of distressed institutions partitioned according to the three peer groups in 2007; similarly, column *Total (2006)* stands for the total number of distressed institutions based on classification in 2006. By *Sub-Total (2005)* we indicate the number of distressed institutions for peer groups in 2005 within partitions of 2006; to obtain the total number of distressed institutions per peer group of 2005 it is needed to sum up across groups in 2006. Source: Authors' own elaborations.

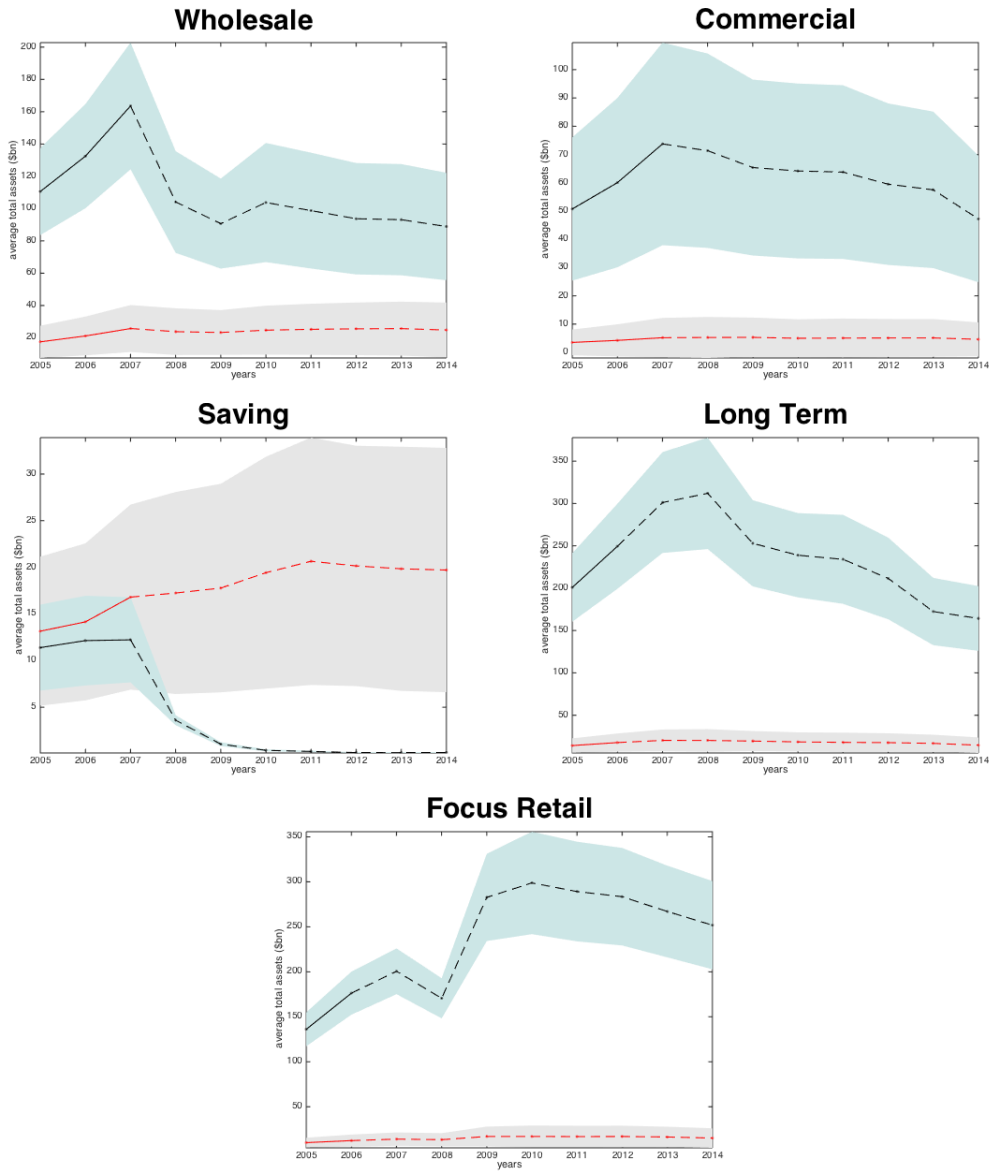
$\downarrow 2006 \backslash 2007 \rightarrow$	Wholesale	Commercial	Saving	Total (2006) $\backslash Sub - Total(2005)$
<b>Wholesale (2005 <math>\downarrow</math>)</b>	<b>67</b>	<b>6</b>	<b>2</b>	<b>75</b>
<i>Wholesale</i>	64	4	2	70
<i>Commercial</i>	1	1	0	2
<i>Saving</i>	1	0	0	1
<i>Long Term</i>	1	0	0	1
<i>Focus Retail</i>	0	1	0	1
<b>Commercial (2005 <math>\downarrow</math>)</b>	<b>4</b>	<b>31</b>	<b>0</b>	<b>35</b>
<i>Wholesale</i>	1	1	0	2
<i>Commercial</i>	3	28	0	31
<i>Saving</i>	0	0	0	0
<i>Long Term</i>	0	2	0	2
<i>Focus Retail</i>	0	0	0	0
<b>Saving (2005 <math>\downarrow</math>)</b>	<b>1</b>	<b>0</b>	<b>51</b>	<b>52</b>
<i>Wholesale</i>	0	0	1	1
<i>Commercial</i>	0	0	0	0
<i>Saving</i>	1	0	50	51
<i>Long Term</i>	0	0	0	0
<i>Focus Retail</i>	0	0	0	0
<b>Long Term (2005 <math>\downarrow</math>)</b>	<b>9</b>	<b>17</b>	<b>4</b>	<b>30</b>
<i>Wholesale</i>	2	3	0	5
<i>Commercial</i>	0	0	0	0
<i>Saving</i>	1	0	0	1
<i>Long Term</i>	6	14	4	24
<i>Focus Retail</i>	0	0	0	0
<b>Focus Retail (2005 <math>\downarrow</math>)</b>	<b>11</b>	<b>1</b>	<b>0</b>	<b>12</b>
<i>Wholesale</i>	0	0	0	0
<i>Commercial</i>	0	0	0	0
<i>Long Term</i>	0	0	0	0
<i>Focus Retail</i>	11	1	0	12
<b>Total (2007)</b>	<b>92</b>	<b>55</b>	<b>57</b>	<b>204</b>

We first focus on the association between distress events and business models. Table 8 presents the distribution of 2008-10 distress events per business model according to annual classifications in the interval 2005-07. For instance, we find that 64 distress events refer to institutions belonging to the *Wholesale* group during the period 2005-07, and similarly we get 50 and 28 distress events for institutions always belonging to the *Saving* and the *Commercial* models in that interval, respectively. The small proportion of distress events over the total number of institutions<sup>27</sup> in the sample should not be taken lightly as they account for a large proportion of total assets (see e.g. Figure 4). For *Wholesale* group, the 64 banks under distress account for \$10.5tn of total assets in 2007, 1/6 of the total of their peers, and average sizes of more than \$160bn, almost 6 times bigger than their peers. Similar results characterize the *Commercial* model, where those 28 distressed institutions reached a total assets coverage of \$2tn in 2007, 1/5 of the whole amount of the group with average sizes up to 15 times those of their peer group members. Different scenario for the 50 distress institutions in the *Saving* group in 2005-07. Their total assets was just \$600bn compared to the \$24.5tn of the whole group in 2007, with average sizes even below the average of their peers (\$12.2bn for the distressed institutions compared with the group average of \$16.8bn in

<sup>27</sup>We recall that the total number of institutions considered in the model for risk assessment is 8526. Institutions that have been always in the same peer group in the interval 2005-07 are: 2355 (*Wholesale*), 2053 (*Commercial*) and 1460 (*Saving*).

2007). Although the majority of distress events are found in wholesale-oriented models, depicting a fragile business model as reported in Roengpitya et al. (2014), Ayadi et al. (2012) and Demyanyk and Hasan (2010), along with the largest share of total assets at stake, distress events are quite well-spread across models, which makes our task of relating specific peer groups to resilience challenging. The remaining 62 distress events are found with institutions switching models, for example 14 *Long Term* institutions in the biennium 2005-06 moving to *Commercial* in 2007, or similarly 11 *Focus Retail* institutions migrating to the *Wholesale* model and so on, with some mix results in terms of relative sizes. The presence of distress events among institutions which exhibit switching patterns (and also the fact that some distress events are related to the two marginal peer groups) motivates our next step of including the switching attitude of institutions into the model to assess the likelihood of distress.

**Figure 4: Total Assets Distribution prior to the Crisis.** Red curves represent the time series of average total assets (in US Billion) for all institutions belonging to each group present prior to the crisis, separately; blue curves stand for distressed institutions only. Dispersion area refers to  $\pm 0.1\sigma$ . Membership to that group refers to the entire interval 2005-07 and is emphasized by continuous lines (period 2005-06 for both Long Term and Focus Retail). Source: Bankscope, Authors' own elaborations.



## 7.1 The Baseline Model of Distress

Table 9 introduce our baseline models to assess the likelihood of distress. The first three specifications regress our dependent variable, i.e. distress events, on each group of regressors separately, i.e. proxies for CAMELS, the macro and the sectoral dimensions, respectively. As reported by Cole and White (2012), DeYoung and Torna (2013) and Jin et al. (2011), among others, balance sheet indicators have been successfully included in the assessment of banks' probability of distress together with macro and sectoral controls to improve predictive information.

For accounting-based measures of banks' business conditions we therefore follow literature and we consider CAMELS proxies (see e.g. Betz et al., 2014; Poghosyan and Čihák, 2009). Table 9 indicates that negative and significant effects are associated to Capital<sup>28</sup>, Cost to Income Ratio, Net Interest margins and liquidity ratios, while positive impacts are found with ROE, Total Securities to Total Assets and Capital Funding Ratio.

ROE is usually interpreted in terms of bank risk taking, so more risky activities seem to imply higher probability of distress. We anticipate that ROE can eventually have a negative impact depending on the peer group, which will support our intuition on **H1** to discriminate risk drivers among business models (details in Section 7.4).

Large liquid positions influence this risk negatively, as higher levels of liquidity might enforce the solvency of the institution (see e.g. Khan et al., 2016). The recent financial crisis posed severe concerns on the funding liquidity risk embedded in balance sheet structures. The maturity transformation related to the funding of long-term mortgages and other securitised exposures using short-term liabilities (Farhi and Tirole, 2012) played a central role in the propagation of distress<sup>29</sup>, where liquidity unbalances facilitated fire-sales practices which mutually enforced contagion and cascade effects (see e.g. Adrian and Shin, 2010; Krishnamurthy, 2010; Mitchell and Pulvino, 2012). Recognizing the importance of funding and liquidity as contributors to the crisis suggests that the dependence of short-term borrowings may have exacerbated instability in the adoption of business models prior to the outbreak of financial markets (Tirole, 2011). Banks usually follow assets & liabilities management (i.e. ALM) procedures to gauge the liquidity (DeYoung and Jang, 2016), interest rates and currency mismatches of their exposures and to assess the buoyancy of funding and liquidity sources. Failure to properly manage these sensitivities can undermine the quality of balance sheet and make the bank more prone to distress.

Less clear is the impact of Capital, as more equity seems to favour safer levels (Berger and Bouwman, 2013) as found in similar models proposed by, for instance, Betz et al. (2014) and Vazquez and Federico (2015), although once considered together with subordinated debts to mimic the regulatory capital (i.e. the Capital Funding Ratio) the sign becomes positive, although poor statistically significant. A caveat of this interpretation might be based on the potential presence of either non-linear or thresholds effects (Estrella et al., 2000); it also suffers from the multicollinearity issue which will motivate the more parsimonious model that will be described below.

Finally, the share of securities to total assets (Total security to Total Assets) resembles the sensitivity to market risk. The relationship with the probability of distress can be ambiguous since securities represent a volatile source of income, although at the same time these assets can be more liquid than, for example, loans. This feature is particularly relevant for risk assessment during the recent crisis since the effects of fire-sales made some institutions more vulnerable (for a deep review on the topic see Shleifer and Vishny, 2011). Our baseline models point to a marginal and positive role for this regressor, although often not statistically significant.

We will stress here the importance of macro and sectoral indicators along with the standard micro ones to improve the risk assessment of financial institutions. At the macro level, higher GDP per capita reduces the risk of distress, so better economic conditions seem to foster financial system resilience, whereas Inflation and FDI Outflows increase the likelihood of institution's distress (Ostry et al., 2012). We note that the coefficient for House Price has a positive and significant sign which probably outlines the role of the mortgage market whose collapse during the recent crisis heavily influenced financial stability (Cole and White, 2012). Odd sign of Unemployment in the *Macro* model disappears when we introduce a more comprehensive list of regressors. Expected

<sup>28</sup>An interesting empirical analysis on the impact of capital can be found in Berger and Bouwman (2013).

<sup>29</sup>Paligorova and Santos (2016) provides empirical evidence on the maturity transformation attitude of wholesale-oriented compared with deposit oriented banks.

results arise when considering the sectoral analysis, with Market Index returns working in the same direction as GDP per capita, coherently with the high level of correlation among these variables as reported in Table 7. Marginal negative effects appear for the Central Government Debt and even for the proportion of domestic financial sector NPLs to total loans (Bank NPLs to Gross Loan). Opposite contribution to institutions' resilience is associated with sovereign debt yields (Gvt. Long-Term Yield) whose dynamics might reflect the country risk appetite by investors. The presence of sovereign instruments in banks' balance sheets often relate to regulatory requirements and the need to fulfil capital constraints based on the amount of risk-weighted assets<sup>30</sup>. Therefore, financial institutions are used to present relevant exposures to sovereign debt, especially domestic issuances, which influence banking practices and impact on the likelihood of being vulnerable to distress from shocks arising in the sovereign debt markets (see e.g. Laeven and Valencia, 2012). Hence, macro-economic and sectoral conditions impact on the likelihood of distress thus supporting the inclusion of macro-prudential policies as a complement to micro-prudential, bank level, tools.

For completeness, we also run a specification with all regressors, even though we note that the large number of variables compared with the amount of observations and distress events does not allow for an accurate and proper econometric investigation. The fifth specification aims to circumscribe this issue by implementing a parsimonious model using half of the regressors. The selection of variables is mainly driven by their coverage within the sample and the significance of the estimates within the single models, but still preserving the representation of all CAMELS dimensions as well as macro and sectoral control variables. Among the individual regressors in this enriched framework we notice that ROA and liquidity (Deposits to Total Funding) foster the resilience of the institutions; an opposite contribution emerges for ROE that suggests that excessive growth of risky investments can cause institutions to be in distress. At the macro level, we confirm the impact of country business cycle approximated by the GDP per capita, while high Inflation and Unemployment levels contribute to deteriorating stability conditions. At the sectoral level, Market Index has the same negative sign as GDP per capita as presented above. The sectoral market index (Sector Index, i.e. market returns of the financial sector) however is positively and significantly related to distress, capturing the effect of the financial sector bubble prior to the crisis. It is worth noting that the pseudo  $R^2$  of the single model regressions are low, while with the inclusion of a more comprehensive list of regressors (models *Selected* and *Benchmark*) we reach values of explained variability in bank probability of distress similar to the ones obtained by Betz et al. (2014) and Vazquez and Federico (2015).

The last model specification (*Benchmark*) takes also into account membership volatility of institutions in switching business models before the onset of the crisis<sup>31</sup>. We aim at testing whether the volatility of business models adopted by institutions can be interpreted as a sign of vulnerability which increases the probability of distress. We observe a statistically significant coefficient of the categorical variable capturing the behaviour of an institution in switching model twice before the crisis (namely Switch Group = 2), while with respect to the *Selected* model the signs and magnitudes of the other regressors are confirmed as well as the significant effects. Note that institutions in the *Long Term* and *Focus Retail* groups all moved to one of the three core peer groups in 2007, which does not necessarily mean that they have substantially modified their business models (see the discussion in Section 6). Hence, *two* switches in the three year period before 2008 identify those institutions that really present unstable peer group membership, thus representing a better proxy for the effective volatility of business model membership.

<sup>30</sup>For instance prior to the crisis banks under Basel regulations usually benefit from lower risk-weights for positions on government bonds instead of loans. For details see e.g. BCBS (2013).

<sup>31</sup>Due to the methodological choice to use average values prior to the crisis for the regressors, we focus on institutions that are present during the entire interval 2005-07 in the models for the assessment of the risk of distress (our first attempt to test for *H2*). The resulting institutions that are discarded, because they belong to very small groups or are even singletons, are just 12 (according to the clustering of 2005). We also remove 24 institutions mainly supranational for which macro and sector regressor values are not comparable.

**Table 9: Distress Assessment.** The first three models refer to specifications within the corresponding group of regressors: CAMELS, Macro and Sector, respectively. Column *All* stands for the model with the entire set of regressors. Column *Selected* represents a parsimonious model where we include about half of the regressors according mainly to data availability and significance of the coefficients in the single specifications of the model. Column *Benchmark* adds the categorical variable *Switch Group* to the *Selected* specification. *Switch Group* assumes value 1 if the institution switches peer group only one time in the period 2005-07; it assumes value 2 if it switches two times, while it is 0 otherwise. For regressors definitions see Section 4.3. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimensions.

	Camels	Macro	Sector	All	Selected	Benchmark
Intercept	−0.740 (0.459)	−2.661*** (0.331)	−2.462 (1.710)	4.005 (4.420)	−0.632 (0.473)	−0.752 (0.496)
<sup>C</sup> Capital	−0.048** (0.023)			0.024 (0.021)	−0.010 (0.013)	−0.010 (0.013)
<sup>C</sup> Capital Funding Ratio	0.036* (0.019)			−0.038** (0.017)		
<sup>A</sup> Roa	−0.017 (0.074)			−0.306*** (0.110)	−0.217** (0.098)	−0.219** (0.097)
<sup>M</sup> Cost to Income Ratio	−0.015*** (0.005)			−0.009* (0.006)		
<sup>M</sup> Roe	0.023*** (0.008)			0.031** (0.014)	0.035*** (0.010)	0.035*** (0.010)
<sup>E</sup> Net Interest Margin	−0.047* (0.026)			0.020** (0.009)	−0.070 (0.052)	−0.063 (0.052)
<sup>E</sup> Interest Expenses to Total Liabilities	0.026 (0.017)			0.086*** (0.031)		
<sup>L</sup> Liquid Assets to Short – Term Funding	−0.004** (0.002)			−0.001 (0.002)		
<sup>L</sup> Deposits to Total Funding	−0.026*** (0.004)			−0.022*** (0.005)	−0.022*** (0.004)	−0.022*** (0.004)
<sup>S</sup> Total Securities to Total Assets	0.010** (0.005)			0.006 (0.005)	0.004 (0.005)	0.005 (0.005)
GDP per capita		−1.062*** (0.129)		−0.223 (0.579)	−0.605*** (0.215)	−0.592*** (0.215)
Inflation		0.560*** (0.083)		0.170 (0.362)	0.415*** (0.080)	0.420*** (0.081)
House Price		0.085*** (0.022)		0.220*** (0.069)		
Unemployment		−0.144*** (0.049)		0.340** (0.151)	0.078* (0.041)	0.070* (0.041)
FDI-Inflows		−0.017 (0.028)		0.142 (0.142)		
FDI-Outflows		0.088*** (0.026)		−0.058 (0.187)		
Central Gvt. Debt			−0.010* (0.006)	−0.015 (0.024)		
Gvt. Long-Term Yield			0.399** (0.166)	−0.795 (0.796)		
Bank NPLs to Gross Loans			−0.201** (0.091)	−0.332*** (0.125)	−0.400*** (0.080)	−0.396*** (0.081)
Credit to Non-Financial Sector			0.003 (0.008)	−0.018* (0.010)		
Market Index			−0.129*** (0.021)	−0.133 (0.102)	−0.073*** (0.019)	−0.075*** (0.019)
Sector Index			0.023 (0.018)	−0.032 (0.091)	0.026** (0.012)	0.025** (0.012)
Stock Traded			−0.004** (0.002)	0.003 (0.003)		
Switch Group = 1				−0.005 (0.256)		0.100 (0.211)
Switch Group = 2				1.166* (0.605)		0.963* (0.543)
Num. obs.	7515	7251	6243	5292	6584	6584
Num. Distress Events	170	179	153	120	137	137
McFadden's Pseudo R <sup>2</sup>	0.042	0.083	0.069	0.242	0.119	0.118
McFadden's Adjusted Pseudo R <sup>2</sup>	0.027	0.073	0.056	0.188	0.097	0.093

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

## 7.2 Volatility of Banks Business Models

As shown in Table 9, an erratic migration dynamics can favour the probability of distress events. Thus, next analysis focuses specifically on the switching behaviour and introduces the transitions between peer groups as regressors to estimate whether the adoption of certain business models or changing peer group membership affect the risk of distress. Once again we rely on a Penalized-likelihood logistic regression to test **H2**, but differently from previous models here we use some proxies to fill missing values in order to better describe the flows of institutions across peer groups<sup>32</sup>. Due to data limitations in the number of observations and distress events, we rely on a simpler version of the *Benchmark* model discussed in Table 9. Since our focus here is the study of the switching dynamics, we prefer to circumscribe the set of CAMELS, macro and sector regressors in a very parsimonious way (similar to the ones presented and discussed also in Table 12). Estimates in Table 10 show that the sign and the magnitude of the coefficients for CAMELS, macro and sectoral models are very similar across models and in line with those presented in Table 9. For the sake of conciseness, we refer to previous comments on the impact of these regressors.

*Model A, B and C* introduce the peer group membership for year 2005, 2006 and 2007, separately. Interestingly, *Long Term* model seems to be safer than the others, while *Commercial* group appears to be affected by the re-organization of 2007 when institutions converged to three core business models. Therefore, in *Model D* and *Model E* we specifically focus on years 2006 and 2007, just before the outbreak of 2008. Estimates among these two models are quite similar and support the intuition that *Commercial* model might have reduced its vulnerability level with respect to the other core groups. This is coherent with the transition flows discussed in Section 6, where we observe a consistent migration of institutions from *Long Term* to *Commercial* in 2007. As reported in *Models A-B*, *Long Term* institutions seem to be less prone to distress and their move into the *Commercial* group makes the latter safer than the other core models. Also, we note that the coefficient for the switching categorical variable (*Switch Group* = 2) is significant and positive as previously seen in Table 9. This is a further confirmation that this very parsimonious framework provides estimates similar to those obtained with a more structured setup. The presence of significant estimates for *Switch Group* is also interesting as motivating a deeper investigation of the actual flows across business models, as briefly outlined above.

*Model F* decomposes the switches into the different combinations of migrations across peer groups between years 2006 and 2007 captured by the interaction of the categorical variables. We observe that changing from the *Long Term* to the *Commercial* peer group reduces the risk of distress, while conversely a migration from *Long Term* into the *Saving* group increases it. Although the first transition seems pretty consistent with the structure of institutions' balance sheets as discussed in Section 6.4, the latter transition from a wholesale-oriented to a deposit-oriented model might be a sign of funding constraints. The limited availability of debt funding due to increased perception of risk of the institution, amplified by low equity (see Adrian and Shin, 2010) could have caused a liquidity shortfall and a sudden reduction of wholesale funding compared to the deposits. Furthermore, transitions into the *Wholesale* group increases the risk of distress. That could depict a process of amplification of wholesale funding that could worsen the level of stable funding leading to potential higher vulnerability at the onset of the crisis when funding liquidity risk impacted the probability to novate funding sources needed to roll over the investments positions.

Finally, *Model G* in the last column replicates the partition of institutions according to the wholesale-oriented classification as presented in literature. Basically, we merge institutions belonging to the wholesale-oriented category (i.e. *Wholesale*, *Commercial*, *Long Term* and *Focus Retail*) and we present a model similar to *Model F*. By collapsing all models with similar funding structure, we aim at verifying whether wholesale-oriented business models are more prone to vulnerability and distress events compared to the deposit-oriented model (see e.g. Ayadi et al., 2012; Roengpitya et al., 2014; Damar et al., 2013). Estimates point to the absence of significant effects for both business models membership and for transitions across them, which seems to support the intuition

<sup>32</sup>For each group we admit the presence of one missing value in the computation of average values for regressors, while for macro and sectoral variables we also consider geographical aggregated proxies to fill missing values; in addition, for distressed institutions we exploit additional data from FDIC to replace missing values. In particular, regressors for the *Commercial* group present several missing values, which heavily reduce the number of distressed institutions once we use the complete case without proxies in the regression.



that traditional broad classifications may be less likely to detect relevant effects and to granularly differentiate the risk among business models. This is an important contribution of the study, since the introduction of a more accurate representation of banking activities better discriminates among institutions and place them into more accurate and suitable peer groups.

**Table 10: Distress Assessment: Transition Analysis.** All the model specifications share the same very parsimonious selection of CAMELS, Macro and Sector regressors. *Model A*, *Model B* and *Model C* add peer group membership of the institutions in year 2005, 2006 and 2007, respectively; *Model D* includes both 2006 and 2007 peer group membership; *Model E* adds to *Model D* the switching categorical variable (*Switch Group*); *Model F* considers the peer group membership in year 2007 plus the transition dynamics across groups between 2006 and 2007. The reference group for these models is the *Wholesale* model in 2006. *Model G* is similar to *Model F* but collapses all wholesale-oriented groups together and uses the *Saving* group as reference one for the categorical variable used for the switching flows. For regressors definitions see Section 4.3. Due to the presence of missing values which affect the representativity for some groups, thus preventing a proper study of the interaction effects across them, we prefer to add some proxies. Therefore, we admit the presence of one missing value in the computation of average values for regressors and we replace missing values for macro and sectoral variables with geographical aggregated proxies; in addition, we also exploit additional data from FDIC for distressed institutions. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimensions.

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Intercept	-0.929** (0.463)	-0.996** (0.454)	-1.131*** (0.423)	-0.759 (0.464)	-0.665 (0.473)	-0.747 (0.469)	-1.404*** (0.449)
<sup>C</sup> Capital	-0.001 (0.010)	0.001 (0.010)	-0.004 (0.010)	0.000 (0.010)	0.001 (0.010)	0.001 (0.010)	-0.003 (0.011)
<sup>A</sup> Roa	-0.258*** (0.070)	-0.257*** (0.070)	-0.271*** (0.071)	-0.261*** (0.069)	-0.258*** (0.069)	-0.262*** (0.067)	-0.273*** (0.072)
<sup>M</sup> Roe	0.042*** (0.009)	0.042*** (0.009)	0.039*** (0.008)	0.042*** (0.008)	0.042*** (0.008)	0.041*** (0.008)	0.039*** (0.008)
<sup>E</sup> Net Interest Margin	-0.056 (0.043)	-0.061 (0.043)	-0.074* (0.039)	-0.057 (0.042)	-0.055 (0.042)	-0.037 (0.041)	-0.083** (0.037)
<sup>L</sup> Deposits to Total Funding	-0.022*** (0.004)	-0.022*** (0.004)	-0.019*** (0.004)	-0.023*** (0.004)	-0.023*** (0.004)	-0.023*** (0.004)	-0.018*** (0.004)
<sup>S</sup> Total Securities to Total Assets	0.004 (0.005)	0.004 (0.005)	0.003 (0.005)	0.003 (0.005)	0.003 (0.005)	0.003 (0.005)	0.004 (0.005)
GDP per capita	-0.201** (0.088)	-0.173** (0.085)	-0.132* (0.077)	-0.201** (0.087)	-0.206** (0.087)	-0.260*** (0.088)	-0.104 (0.074)
Inflation	0.445*** (0.098)	0.425*** (0.094)	0.353*** (0.085)	0.437*** (0.096)	0.442*** (0.097)	0.484*** (0.098)	0.353*** (0.083)
Gvt. Long-Term Yield	-0.081 (0.093)	-0.073 (0.091)	-0.022 (0.080)	-0.086 (0.093)	-0.095 (0.095)	-0.117 (0.095)	-0.010 (0.077)
Market Index	-0.078*** (0.009)	-0.079*** (0.009)	-0.071*** (0.008)	-0.078*** (0.009)	-0.079*** (0.009)	-0.077*** (0.009)	-0.073*** (0.008)
Group (2005): Commercial	0.597** (0.247)						
Group (2005): Saving	0.034 (0.213)						
Group (2005): Long Term	-0.818*** (0.264)						
Group (2005): Focus Retail	0.582* (0.346)						
Group (2006): Commercial		0.650*** (0.252)		1.379*** (0.390)	1.382*** (0.398)		
Group (2006): Saving		0.019 (0.216)		0.283 (0.408)	0.129 (0.443)		
Group (2006): Long Term		-0.654*** (0.252)		-0.014 (0.352)	0.129 (0.427)		
Group (2006): Focus Retail		0.534 (0.353)		0.470 (0.354)	0.708 (0.434)		
Group (2007): Commercial			-0.384* (0.199)	-0.861** (0.354)	-0.817** (0.358)	0.871* (0.467)	
Group (2007): Saving			-0.102 (0.202)	-0.315 (0.396)	-0.170 (0.430)	-0.773 (0.680)	
Group (2007): Wholesale-oriented							-0.409 (0.852)
Switch Group = 1				(0.396)	-0.297 (0.286)		
Switch Group = 2					0.962* (0.567)		
Group (2006) Commercial $\mapsto$ Group (2007) Wholesale						1.884*** (0.634)	
Group (2006) Saving $\mapsto$ Group (2007) Wholesale						0.746 (0.884)	
Group (2006) Long Term $\mapsto$ Group (2007) Wholesale						0.689* (0.404)	
Group (2006) Focus Retail $\mapsto$ Group (2007) Wholesale						0.507 (0.367)	
Group (2006) Commercial $\mapsto$ Group (2007) Commercial						-0.224 (0.494)	
Group (2006) Saving $\mapsto$ Group (2007) Commercial						-1.747 (1.518)	
Group (2006) Long Term $\mapsto$ Group (2007) Commercial						-2.104*** (0.517)	
Group (2006) Focus Retail $\mapsto$ Group (2007) Commercial						1.441 (1.134)	
Group (2006) Commercial $\mapsto$ Group (2007) Saving						0.123 (1.688)	
Group (2006) Saving $\mapsto$ Group (2007) Saving						0.788 (0.678)	
Group (2006) Long Term $\mapsto$ Group (2007) Saving						2.351*** (0.874)	
Group (2006) Wholesale-oriented $\mapsto$ Group (2007) Saving							-0.124 (0.438)
Group (2006) Wholesale-oriented $\mapsto$ Group (2007) Wholesale-oriented							0.351 0.841
Num. obs.	6795	6795	6795	6795	6795	6795	6795
Num. Distress Events	183	183	183	183	183	183	183
McFadden's Pseudo R <sup>2</sup>	0.111	0.108	0.101	0.109	0.110	0.122	0.099
McFadden's Adjusted Pseudo R <sup>2</sup>	0.092	0.089	0.084	0.087	0.085	0.092	0.080

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1

### 7.3 Switch Analysis

In this Section we attempt to shed a light on the main driving forces responsible for institutions' transitions across business models in line with *H2bis*. Although our main goal is the assessment of distress, our findings suggest that business model volatility plays an important role to the resilience of financial institutions. Therefore, an attempt to understand the main driving forces behind these switching patterns can assist our study of business model volatility and distress<sup>33</sup>. In order to concentrate our investigation on the wave of financial turmoil originated from the 2007-08 crisis, we circumscribe this analysis to both switches from 2005 to 2006 and from 2006 to 2007. We propose a basic framework similar to the one applied to study the dynamics of distresses to investigate the reasons for switching business models. We rely on a Firth's Penalized-likelihood logistic regression model (Firth, 1993) to explain whether a financial institution switches model ( $Y = 1$ ) or stays in the same group ( $Y = 0$ ). We regress switches against a list of regressors computed as averages over the three years before the event. Since we aim to present a simple representation of the switching dynamics prior to the crisis, we consider together both switches from 2005 to 2006 and those from 2006 to 2007 with the corresponding regressors calculated over the 3-years interval before the switch to ameliorate potential endogeneity problems.

We consider three potential determinant classes: *i*) the characteristics of the institution, in terms of size, leverage, and maturity funding and liquidity mismatches to assess whether main internal factors influence the likelihood to modify the respective business model; *ii*) macroeconomic and sectoral conditions, to assess the impact of external factors; *iii*) the membership to a certain business model, to understand whether the adoption of some business models could be considered a transient state.

To control for possible size effects, we consider whether the amount of Total Assets impacts on the switching dynamics. We recall that the vectors of measures used to compute the cosine similarities are normalized by institutions' respective total assets, hence the identification of peer groups is mainly driven by similarities among balance sheet ratios. We introduce Equity/Total Assets as a measure of capitalization and leverage, Deposit and ST Funding/Total Funding to approximate the weight of traditional funding sources, and finally Liquid Assets/Deposit and ST Funding as a measure of net liquidity that capture the amount of maturity transformation of the institutions. Overall these basic variables aim to concisely depict banks in terms of their level of size, leverage and funding-liquidity mismatches, which we assume are among the most active forces affecting migrations across business models. The GDP per capita and Market Index are introduced to control how economic and financial conditions influence institutions resilience in the same group over time, or alternatively, determine the tendency to change model. Finally, we investigate how initial business model features affect financial institutions' likelihood to switch model in time  $t+1$  by introducing the model they adopt at time  $t$ .

Since peer group membership of an institution relates to the multi-dimensional distance between itself and the centroid of the respective group compared to other centroids, we expect that institutions that are farther from the core of their group in a certain year are more likely to switch to a different model in the next period. We basically claim that the more peripheral the institution is within its group in terms of business model, and therefore potentially closer to other models, the higher its probability to switch. To measure this distance, we introduce the *Group-Score* (hereinafter G-Score), representing the sum of the distances between each observation from the mean of each measure, within the same group, used to compute the cosine similarities. Moreover, we take into account the dispersion of each variable standardising every distance by the respective standard deviation. Hence, any deviation from a well concentrated measure is penalised more than a departure from a dispersed variable. The standardised score for bank  $i$  is computed within the corresponding peer group as:  $\sum_{j=1}^N |(X_{ij} - \mu_j)|/\sigma_j$ , where  $N$  is the number of variables used to compute the cosine similarities (i.e.  $N = 29$  balance sheet measures),  $X_{ij}$  refers to the obser-

<sup>33</sup>The study of the switching dynamics and of the reasons behind changes in business models is simply outlined in this Section. A complete investigation goes beyond the scope of the paper and is, therefore, left for future research. We stress that among possible causes, not considered in our framework, regulatory policies might influence banks to vary their business models as well as market-based dimensions that affect the assets and liabilities management of the bank. We remark that endogeneity represents a serious issue when accounting-based bank-level variables are included in the model.

vation of bank  $i$  for measure  $j$ , while  $\mu_j$  and  $\sigma_j$  stand for the mean and the standard deviation of measure  $j$ , respectively. Furthermore, since institutions have missing values in some of the  $N$  variables preventing the computation of the distances in these cases, we decide to further divide the score of institution  $i$  by the number of its non-missing variables among the  $N$  fields to enhance comparability.

**Table 11: Transition Models prior to the Crisis.** The first three columns refer to estimates computed for each peer group separately. We consider pooled observations, that means we join together the transitions between 2005 and 2006 with those from 2006 to 2007. Peer group names reported in columns' names refer to the membership before the transitions. For instance *Wholesale* refers to those institutions which were in that group in 2005, whose regressors are computed on average values on the interval 2003-05, and to those institutions which were *Wholesale* in 2006, whose corresponding regressors are averaged over the interval 2004-06. *Model 1* and *Model 2* include all the institutions regardless their peer group, while models with asterisks refer to the exclusion of institutions belonging to *Long Term* and *Focus Retail* groups in the biennium 2005-07. Variable *Group* stands for the membership in  $t-1$  to a specific business model (the reference level is the *Wholesale* group). Variable *G-Score* refers to the year prior to the transition. Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Model 1	Model 1*	Model 2	Model 2*
Intercept	-2.264*** (0.473)	-6.021*** (0.749)	-2.890*** (0.950)	-1.876*** (0.208)	-3.813*** (0.309)	-3.766*** (0.259)	-3.997*** (0.320)
Total Assets	0.179 (0.538)	1.338* (0.706)	-0.801 (1.557)	0.124 (0.282)	0.458 (0.396)	0.445 (0.342)	0.500 (0.395)
Equity/Total Assets	-0.008 (0.006)	-0.014 (0.013)	0.017 (0.012)	-0.021*** (0.004)	-0.009** (0.004)	-0.012*** (0.004)	-0.009** (0.004)
Deposits ad ST Funding/Total Funding	0.639 (0.431)	1.970*** (0.733)	-1.803* (0.947)	-0.830*** (0.195)	0.158 (0.296)	0.227 (0.237)	0.313 (0.304)
Liquidity Assets/Deposits and ST Funding	0.004*** (0.001)	0.015*** (0.003)	0.009** (0.004)	0.002*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
GDP per capita	-0.087** (0.040)	0.245*** (0.055)	0.105** (0.044)	0.153*** (0.014)	0.006 (0.023)	0.104*** (0.019)	0.033 (0.024)
Market Index	-0.024*** (0.005)	0.003 (0.008)	0.017*** (0.004)	-0.014*** (0.002)	0.001 (0.003)	-0.011*** (0.003)	-0.001 (0.003)
G-Score (t-1)	0.055 (0.283)	1.608*** (0.199)	1.134*** (0.278)	1.357*** (0.098)	1.552*** (0.118)	1.394*** (0.109)	1.551*** (0.118)
Group(t-1): Commercial						0.369*** (0.109)	0.269** (0.111)
Group(t-1): Saving						-0.178 (0.111)	-0.167 (0.112)
Group(t-1): Long Term						3.298*** (0.115)	
Group(t-1): Focus Retail						2.962*** (0.117)	
Num. obs.	3291	2575	2778	9862	8644	9862	8644
Num. Switch	255	217	166	1405	638	1405	638
McFadden's Pseudo R <sup>2</sup>	0.021	0.156	0.073	0.052	0.049	0.274	0.050
McFadden's Adjusted Pseudo R <sup>2</sup>	0.010	0.144	0.058	0.049	0.045	0.271	0.045
LR Test	71.566***	257.941***	119.887***	461.577***	262.764***	2261.887***	277.221***

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 11 exhibits estimates for the transition dynamics from 2005 to 2006 and from 2006 to 2007 together. *Model 1* and *Model 2* include all the institutions regardless their peer group, while models with asterisks exclude institutions belonging to *Long Term* and *Focus Retail* groups in the biennium 2005-07. We also report estimates of models including only institutions belonging to one of the main three core models<sup>34</sup>.

As expected, G-Scores have a significant and positive effect in all specifications that consider the whole sample. Hence, institutions that are more peripheral in a certain group are more prone to switch to a different one. However, when inspecting this effect within each group we observe that G-Scores is positive in all core groups but only statistically significant for *Commercial* and *Saving* banks. Although those results must be taken with caution due to the small sample size, the lower volatility of this variable within the *Wholesale* group compared to other groups can justify the low and not significant value of its coefficient. It also raises the issue of the *dark side of diversification* (Oordt, 2013) reviewed in Section 2 where the high level of diversification within a peer group makes the whole group very similar with dangerous systemic consequences.

Total Assets is not significant in any specification of the model, discarding a size effect on the likelihood of business model transitions. One might argue that large institutions are more

<sup>34</sup>We omit models that include only *Long Term* or *Focus Retail* institutions due to the lack of a sufficient amount of observations. Note that the impact of those institutions is however captured by the comparison between *Model 1* and *Model 2* and their relative asterisk versions.

flexible in the choice of their balance sheet structures, relying on a much more diversified spectrum of possible investments and a vast set of funding sources, thus being facilitated whenever they plan to change their business model; however, one might also notice that these institutions suffer from more regulatory constraints and that large banks reflect the interests of a wider perimeter of stakeholders, making the switching process potentially slower<sup>35</sup>. Our results suggest that being larger in terms of total assets do not facilitate a switching dynamics at least before the outbreak of 2007-08. A positive and slightly significant effect is observed only for the *Commercial* model.

The negative and significant effect of Equity/Total Assets is well established in literature and confirmed by our models. On the governance side, large equity size usually suggests higher shareholder control over the activities of the institution that usually limit the ability of managers to change business models (Dell’Arriccia and Marquez, 2010; Iannotta et al., 2007; Manzaneeque et al., 2015; Mori and Ikeda, 2015). On the market side, large amount of equity, probably amplified by assets prices boom in the pre-crisis period, would facilitate the search for funding and therefore the ability of the institution to meet their strategy when actively managing their balance sheets (Adrian and Shin, 2010). Empirical findings show that wholesale-oriented banks tend to have an advantage on that over the deposit-oriented banks (Damar et al., 2013) that is confirmed by the sign of the coefficient within each model (even though the effect is not significant).

Deposit and ST Funding/Total Funding captures the short-term debt that is heavily driven by customer deposits. The effect is not obvious when looking at the whole set as it features a negative and significant sign in *Model 1* whereas positive (but not significant) in all the other specifications that consider all the institutions regardless of the peer group. One could argue that this could be the result of a non linear effect similar to that on non-deposit funding discussed in Demirgüç-Kunt and Huizinga (2010), DeYoung and Torna (2013), and Vazquez and Federico (2015). However, when considering it within each business model the effect of short-term debt becomes very clear. Our findings show negative and significant impact within the *Saving* model (e.g. higher short-term funding, mainly deposit, amplifies the current orientation towards deposit-oriented model reducing the likelihood of a switch), while the effect is positive for wholesale-oriented institutions (e.g. higher short-term debt would signal a transformation of the funding towards a deposit-oriented model).

The impact of net liquidity, approximated by Liquid Assets/Deposit and ST Funding, captures the level of maturity transformation of banks, ranging from very low values for traditional models like the *Saving*, to higher levels observed on wholesale-oriented models. The effect is positive and consistent across specifications, both within and across business models, supporting the idea that liquidity provides higher flexibility for quick adjustments of business strategy, regardless of the business model itself, and therefore promoting business model volatility. This effect appears a bit more pronounced withing *Commercial* institutions that would benefit the most, among the wholesale-oriented models, from a better assets diversification.

The impact of macro-economic conditions is significant and the sign of the coefficient is in general positive, with the exception of the *Wholesale* model where the coefficient is significant and negative. Hence, for institutions that are located in countries that exhibit better GDP per capita patterns the change to alternative business models looks easier. Investment opportunities that flourish during economic and financial booms would promote higher diversification, thus having a positive impact on model volatility for *Saving* and *Commercial* models (that do not diversify much on the assets side), while having a negative impact on *Wholesale* banks as it exacerbates their business characteristics. This effect is also supported by the impact of Market Index.

Finally, Table 11 provides additional insights on the likelihood of each business model to switch<sup>36</sup>. *Model 2* features the higher explanatory power across all specifications, suggesting that business model classification improves the quality of the predictions for the whole sample. The coefficient for the categorical variable Group (t-1) when it is equal to *Commercial* (*Group(t-1): Commercial*) is positive and significant, thus suggesting that this business model tends to be more

<sup>35</sup>The extent to which various stakeholders affect bank business decisions as well as the influence of bank portfolio compositions on the size and activities of the institution are discussed from different perspectives e.g. in Benston et al. (1995), Chava et al. (2009), John et al. (2008), Kishan and Opiela (2000), Laeven and Levine (2009), and Milbourn et al. (1999), among others.

<sup>36</sup>Some switching institutions cannot be considered in the models presented in Table 11 due to missing values in the computation of regressors values. In Appendix E we provide additional robustness tests by enlarging the set of switching institutions that can be considered in the estimation model. We anticipate that those results point to a very similar picture as the one shown in Table 11.

volatile than the *Wholesale* one (used as reference group). The *Commercial* model practically sits in between the other two core models, inheriting assets side characteristics from the *Saving* and funding ones from the *Wholesale*. This hybrid feature facilitates the transition to either one of the two models. The last two realization of the categorical variable are expected to be highly significant and positive due to the volatile nature of those models. *Model 2\** focuses on the three core models showing estimates very similar to the ones presented in *Model 2*, validating our findings when restricting the sample to the three core groups only.

## 7.4 Risk Drivers of Banks Business Models

The information of the true business model adopted by banks represents a valuable opportunity for regulators to investigate institutions' characteristics and vulnerability to distress. This would support targeted intervention and more accurate risk monitoring. To disentangle the impact of being in a particular peer group in the pursuit of finally testing **H1**, we analyse each business model separately by partitioning the sample according to stable membership along the interval 2005-07. Table 12 focuses on a restricted case where we discard institutions that switched business models prior to the crisis. This configuration aims to identify risk drivers that are not biased by the impact of switching dynamics. Since the number of observations and, in particular, distress events under each peer group is modest, we need to consider a Penalized-likelihood logistic regression with very few regressors<sup>37</sup>. We propose a simple framework with common CAMELS variables and two basic macro regressors (GDP per capita and Inflation) along with two sectoral variables which capture market dynamics (Government LT Yield and Market Index return). The choice of these regressors is in line to the ones presented in Table 10 producing estimates for the entire set of no-switching institutions similar to those discussed in the *Benchmark* model of Table 9. This supports the selection of these regressors for the reference model used to create specifications for each business model. Note that although confined to specific peer groups membership, these configurations share the same setup of previous models to enhance comparability. In addition, we add the amount of institutions' total assets as regressor. In fact, as highlighted in Section 6.3, size relative to their peers seems to matter for the event of distress during the recent financial crisis at least for the wholesale-oriented institutions.

Table 12 confirms that dominant assets size relative to their peer group members exacerbates the likelihood of distress only on wholesale-oriented models, whereas it has no significant impact for the deposit-oriented group. Estimates indicate that *Wholesale* and *Saving* models present quite different risk drivers in terms of CAMELS variables: ROE impacts positively on the likelihood of distress for *Wholesale* institutions and negatively for *Saving* ones (although the coefficient magnitude is quite small), while ROA exhibits an opposite pattern and capital has a negative sign for *Saving* but does not show significant effects for *Wholesale*. Thus, our results seem to support **H1** by emphasizing different forces affecting institutions' probability of distress under wholesale-oriented or deposit-oriented models. The only exception is on the proxy for liquidity that is negative and consistent across model specifications. This result provides supporting evidence of recent debates that stress the impact of liquidity to the recent financial crisis (Tirole, 2011) that we find independent of the adopted business model.

On the Capital, we recall that *Saving* institutions are less capitalized compared to *Wholesale* and would definitely benefit from a boost of capital to enhance stability (Vazquez and Federico, 2015; Berger and Bouwman, 2013). We might also advance the explanation on the effect of ROA<sup>38</sup>, by comparing the level of diversification on the assets side between deposit-oriented and wholesale-oriented groups, more precisely the differences between a loan-focused versus a well diversified investment strategy. Institutions within the *Saving* model cannot fully exploit a wide spectrum

<sup>37</sup>We omit model specifications for the *Commercial*, *Long Term* and *Focus Retail* only due to the limitation of distress events. In Appendix E we report an alternative specification of the models presented in Table 12 where we use proxies to fill missing values, thus enlarging the sample size. Results still show very similar effects. To enhance comparability, the robustness checks in Appendix E also consider the other models omitted in Table 12.

<sup>38</sup>The assessment of Asset quality within the CAMELS framework usually includes the relationship between loans losses and reserves for impaired loans as an indicator of the quality of the assets side. Balance sheet items for the period prior to the crisis do not allow a proper computation of this relationship. Hence we rely on ROA due to its high coverage in the sample. Basically, this ratio has embedded in the numerator (net income) the contribution of loans losses, while the denominator (total assets) reflects the presence of the corresponding reserves.



of investment choices to boost their returns on assets (see Chiorazzo et al. (2008) for evidence in the EU). This is due to a limited range of instruments in the assets side (recall their investment strategy mainly focused on commercial/corporate loans), which probably impacts on their ability to select risk-adjusted profitable investments compared with *Wholesale* institutions. Empirical evidence shows that deposit funded banks tended to extend more credit lines to firms on longer maturities (Paligorova and Santos, 2016) due to their stable funding (Ivashina and Scharfstein, 2010; Kashyap et al., 2002), that although promoting profitability would expose them to the corporate credit market that was affected by the sub-prime mortgage crisis. On ROE, the opposite effect would be related to the structure of the liabilities side of the institutions. *Wholesale* model may already present higher exposure on profitable and risky assets, and further investment decisions aimed to improve the level of ROE can worsen the sustainability of their activities (Stiroh, 2004) compared with *Saving* model. This effect is in line with the findings of Demirgüç-Kunt and Huizinga (2010) on the nonlinear effect of the growth non-deposit funding and non-income investments to distress. *Wholesale* institutions have higher exposures to interbank debts, and further investments are indeed likely to involve an increase on this type of leverage, which in turn deteriorates the resilience of their business model and, eventually, exacerbates their risk of distress (Vazquez and Federico, 2015); *Saving* institutions on the other side are more characterized by stable funding, namely deposits, so the mix of funding that they can use for investment purposes is less prone to suffer from financial market instability (Beltratti and Stultz, 2012). The latter phenomenon can imply contagion dynamics that are related to interbanking exposures through which contagion may actually propagate (Krause and Giansante, 2012). It is worth recalling that the dependence on interbanking positions is a specific feature of the wholesale-oriented business models, while during the current crisis customer deposits, dominant funding source in deposit-oriented institutions, were not particularly affected by bank runs triggered by the lack of confidence in banks quality (for discussions see e.g. Gorton, 2010). Furthermore, the interconnectivity arisen from interbank exposures might have determined the need to redefine bilateral positions during the outbreak of financial markets, to compensate for the increasing perception of counterpart risk, resulting in more volatile balance sheet compositions. This might have also influenced the reallocation of investments on the assets side, due to constraints on funding sources which reciprocally affected fire-sales dynamics (Anand et al., 2013; Georg, 2013).

Furthermore, similar to the previous case shown in Table 10 we add a specification which includes all the wholesale-oriented institutions in the same group (last column is circumscribed to the merge between *Wholesale* and *Commercial* models only). Although the main discussion above compares the two extreme business models, *Wholesale* and *Saving*, which is in line with the current debates on business models (see Section 2), this further analysis helps validating comparisons between wholesale-oriented models and the *Saving* model, due to the inclusion in particular of *Commercial* institutions. All three specifications of wholesale models proposed are consistent and estimates reinforce the interpretation that wholesale-oriented institutions and deposit-oriented institutions present different risk drivers.

To summarize, our analysis confirms the importance and the sign of CAMELS dimensions in explaining the likelihood of distress of financial institutions during the recent crisis and provides a solid ground for taking the true banks business models into consideration for a more accurate risk assessment and monitoring. Two additional dimensions emerge in this framework: the characteristics of the business model adopted by institutions and the volatility of that decision over time. For the first dimension, CAMELS measures along with macro and sectoral features contribute differently, sometimes with opposite sign, to the likelihood of distress among institutions adopting a different business model with the exception of liquidity. For those institutions who tend to switch models very often, identifying the second dimension of the problem, we observe that business models instability exacerbates vulnerability especially when moving across wholesale-oriented and deposit-oriented model categories. As for the probability of distress, their switching decision is also driven by both micro and macro/sectoral factors with sometimes opposite sign depending of the business model. Our findings suggest that bank supervisor would definitely benefit from monitoring these true business model features for a more accurate and targeted intervention in stabilizing the banking sector.



**Table 12: Distress Assessment within Peer Groups.** Columns *Wholesale* and *Saving* refer to observations for institutions that do not change peer group and that belong to that specific business model indicated in the name of the column. Column *Wholesale-oriented* refers in particular to institutions belonging to groups Wholesale, Commercial, Long Term and Focus Retail, admitting for transitions across these groups and never being in the Saving group in the period 2005-07 (in the case of Long Term and Focus Retail the interval that is considered is 2005-06). Column *Wholesale-oriented (Restricted)* is circumscribed to Wholesale and Commercial models. Column *All* includes all institutions. For regressors definitions see Section 4.3. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimensions. Total Assets are in USD Trillion.

	All	Wholesale	Saving	Wholesale-oriented	Wholesale-oriented (Restricted)
Intercept	-2.282*** (0.537)	-1.130 (1.108)	4.942*** (1.764)	-2.564*** (0.628)	-1.815** (0.925)
<sup>C</sup> Capital	-0.005 (0.015)	0.010 (0.013)	-0.138** (0.055)	0.006 (0.013)	0.018 (0.012)
<sup>A</sup> Roa	-0.092 (0.103)	-0.175* (0.094)	0.678** (0.286)	-0.240** (0.104)	-0.252** (0.105)
<sup>M</sup> Roe	0.025*** (0.009)	0.044*** (0.015)	-0.042** (0.018)	0.038*** (0.011)	0.061*** (0.015)
<sup>E</sup> Net Interest Margin	-0.075* (0.040)	0.012 (0.009)	0.169 (0.166)	-0.082** (0.036)	-0.025 (0.049)
<sup>L</sup> Deposits to Total Funding	-0.018*** (0.004)	-0.020*** (0.006)	-0.066*** (0.015)	-0.017*** (0.005)	-0.016*** (0.006)
<sup>S</sup> Total Securities to Total Assets	-0.007 (0.006)	0.004 (0.008)	-0.003 (0.012)	0.003 (0.007)	0.004 (0.008)
GDP per capita	0.012 (0.107)	-0.144 (0.206)	-3.061*** (1.048)	0.082 (0.100)	-0.043 (0.171)
Inflation	0.196 (0.130)	0.203 (0.220)	-0.150 (0.486)	0.141 (0.135)	0.319 (0.196)
Gvt. Long-Term Yield	0.248* (0.136)	-0.275 (0.287)	0.852* (0.483)	0.050 (0.170)	-0.242 (0.249)
Market Index	-0.085*** (0.013)	-0.021 (0.022)	0.088 (0.062)	-0.050*** (0.013)	-0.057*** (0.017)
Total Assets	2.052*** (0.342)	1.606*** (0.476)	0.029 (1.450)	1.984*** (0.317)	1.531*** (0.395)
Num. obs.	5780	1867	1231	4870	3450
Num. Distress Events	126	46	46	92	59
McFadden's Pseudo R <sup>2</sup>	0.115	0.137	0.180	0.107	0.093
McFadden's Adjusted Pseudo R <sup>2</sup>	0.093	0.067	0.111	0.077	0.048

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

## 8 Conclusions

The study investigates the impact of the true banking business model on vulnerability and likelihood of distress of a large set of banks globally distributed that includes more than 11,000 institutions representing more than 180 countries. Specifically, we test two main hypothesis. The first is related to the impact of the business model adopted by the institutions to their probability of distress. Although empirical evidence suggests that wholesale-oriented models are more prone to distress due to low levels of stable funding, the recent financial crisis exposed the fragility of all business models. We are therefore interested in identifying institutions risk drivers as function of the main business models. The second hypothesis is related to the volatility of business models. Institutions might alter their funding/investment strategy over time to boost resilience and returns. By targeting those institutions that switched business model prior to the crisis, we test whether this switch behaviour is a sign of instability that would lead to distress.

We propose a classification approach specifically designed for large, complex and sparse data sets, aimed to expose banks true business models. The approach employs the cosine similarity measure to compare pairs of institutions and a hierarchical clustering method, the *Louvain* algorithm, very popular in complex science, to define the best configuration of peer groups according to a modularity maximization. Our validation tests confirm that the resulting groups are statistically different, showing a superior clustering fit than the direct specialization well-adopted in banking literature. We also show comparable results between *Louvain* and Ward, i.e the mainstream method in literature, as hierarchical clustering techniques, although the methodological

advantage of dealing with complex and sparse data inputs provided by *Louvain* makes the latter a more suitable approach to deal with our sample.

We find seven well-defined peer groups resembling the three main model categories very stressed in literature: wholesale-oriented, deposit-oriented and investment-oriented groups. Four models fall into the first category, named *Wholesale*, *Commercial*, *Long Term* and *Focus Retail*. Although all these four models are characterized by a relevant wholesale funding, the first model dominates on this dimension while having a well diversified investment strategy; the second still presents high level for wholesale funding although it specializes in commercial/corporate loan investments; the third is well-diversified in the assets side as the *Wholesale* but shows the largest portion of long term funding of all; the last exhibits a dominant position of retail loans investments. We also find two models falling into the deposit-oriented category, namely *Saving* and *Diversified Retail*. Both are characterized by dominant deposit funding. However, only *Saving* model was available throughout the whole period, whereas *Diversified Retail* emerges from 2008 onwards and is composed mainly by German and Swiss saving and cooperative banks that migrated from the two main wholesale-oriented models right at the peak of the financial crisis. Finally, only one model falls into the investment-oriented category, named *Investment*, and represents a non-deposit funding model which is very active in the interbank market as the largest net borrowing institutions with a large exposure to non-income investments. They appear to be huge institutions, among the largest international broker dealers, with asset size almost ten times the size of their peer group competitors. The *Investment* model is only available in 2012 and 2014 when the majority of those institutions migrated from the *Wholesale* model. We also note that direct specialization is a poor indicator of true business model, especially when comparing institutions in a global contest.

We develop a comprehensive and globally distributed distress event list by merging events of bankruptcy and liquidation, defaults, distressed merges and public bailouts from many different sources globally distributed. Statistics of the distress events per business classification show that all business models were affected by distress events, with dominance on the wholesale-oriented models both in numbers (due to the higher popularity of these models over the deposit-oriented ones in the pre-crisis period) and in average total assets sizes. However, we also find a substantial number of distress events among institutions with a very unstable business model, i.e. those who tend to switch peer group quite often prior to the crisis, which motivates a deeper investigation.

As expected, CAMELS dimensions are almost all significant in explaining banks' risk of distress, with negative contributions from capital, cost to income ratio, net interest margins and liquidity ratios, and positive from ROE, total securities to total assets and capital funding ratio. Macro and sectoral characteristics also affect banks' resilience, with GDP per capita and market index boosting stability whereas FDI-outflows, unemployment, house price and government debt yield weakening institutions' resilience. Instability of membership to business models over time has a statistically significant and positive impact to the likelihood of distress too. More specifically, we find out that those wholesale-oriented institutions coming from the *Long Term* model were more resilient to distress, showing that higher long term debt can facilitate banks stability during market turmoil. However, a migration from *Long Term* to a deposit-oriented model right before the onset of the financial crisis had exacerbated vulnerability and risk of distress. This depicts institutions with potential funding restrictions in 2007 as sign of perceived vulnerability that resulted on a reduction of wholesale debt compared to customer deposits. Finally, those institutions that moved to *Wholesale* model in 2007, exposing themselves to more unstable funding sources, were exposed to more risk of distress.

The second set of tests focuses on exposing the main driving forces responsible for institutions transitions across business models. We observe that institutions switching decision is driven by both micro and macro/sectoral factors. The similarity with their peer group members improve business model stability, meaning that peripheral institutions within their peer group tend to be more prone to change business model, and ultimately being more vulnerable to distress. Higher capitalization, providing higher shareholder control and facilitating the search for funding sources, also improves stability of membership. Opposite and consistent effect for liquidity across peer groups, providing higher flexibility for adjustments of business models. However, other factors affect switching behavior differently depending on the business model. For example short-term funding negatively affect switches for deposit-oriented institutions, due to the large impact of deposit for this business model, while positively affecting wholesale-oriented institutions, which

conversely rely more on unstable funding sources. The inclusion of business models membership improves the explanatory power of our models, suggesting that *Commercial* institutions tends to be more prone to migrate into the other groups due to their similarity to both *Wholesale* and *Saving* on the liabilities and assets side, respectively.

Last set of tests, performed by partitioning the sample according to peer group memberships that were consistent over the period 2005-07, aims at considering only institutions that maintained the same business model prior to the crisis. By observing the contribution of the main banks' characteristics within each business model, we confirm that relative size (in terms of total assets) is a driving force only among wholesale-oriented models, suggesting that being more exposed to a wholesale-oriented activities increases vulnerability. This is not the case within the deposit-oriented group, where relative size does not play a role. We also compare wholesale-oriented vs. deposit-oriented institutions noting opposite patterns for CAMELS proxies contributions to the risk of distress, such as: a negative impact of ROE for *Saving* and positive for *Wholesale* capturing the impact of leverage on stable versus unstable funding models respectively; opposite effects for ROA exposing the impact on higher asset returns on focused versus well diversified investments respectively; significant and negative sign for capital only for the *Saving* group, supporting the idea that capital improved stability mainly for those institutions heavily funded by customer deposits.

To conclude, we provide evidence of how different business models affect institutions' vulnerability and likelihood of distress at a global level, where the one-rule-fits-all approach for monitoring and risk assessment could be dramatically misleading compared to a targeted and bespoke approach designed for business models true classifications. The exception we find is on the impact of liquidity that appears exogenous to business model choice<sup>39</sup>. Regulators will definitely benefit from this analysis as, on top of the CAMELS proxies combined with macro and sectoral info that are currently used to assess the likelihood of banks to distress, two additional information emerge from this study as relevant for risk assessment: the characteristics of the business model adopted by institutions in relations to their own sensitivity to distress and the volatility of that decision over time. Direct extensions to this study may go to the direction of extending the sample period by considering earlier years, with quarterly intervals to improve the estimation procedure and to allow for an early-warning setup. Of course, this would require a more parsimonious set of institutions' characteristics to overcome comparability issues. The vulnerability of banks could also be assessed by other indicators well-adopted in literature, such as distance to defaults (z-score), SRISK, MES, DeltaCOVAR, or DIP, which can be easily included in our framework, as well as more focused macro and sector indicators designed for specific distress events within the geographical vs. business model space.

---

<sup>39</sup>The new adopted Liquidity Coverage Ratio and Net Stable Funding Ratio introduced by Basel III are meant to tackle this issue (BCBS, 2011b).

## References

- Adrian, T. and H. S. Shin (2010). “Liquidity and leverage”. In: *Journal of Financial Intermediation* 19.3, pp. 418–437.
- Anand, K., P. Gai, S. Kapadia, S. Brennan, and M. Willison (2013). “A network model of financial system resilience”. In: *Journal of Economic Behavior & Organization* 85, pp. 219–235.
- Ayadi, R., E. Arbak, and P. W. De Groen (2011). “Business models in European banking: A pre-and post-crisis screening”. In: *Center for European Policy Studies*.
- (2012). “Regulation of European banks and business models: towards a new paradigm?” In: *Centre for European Policy Studies*.
- Ayadi, R. and P. W. De Groen (2015). “Banking business models 2015 Europe”. In: *Alphonse and Dorimène Desjardins International Institute for Cooperatives and International Research Centre on Cooperative Finance*.
- BCBS (2010). “An Assessment of the Long-Term Economic Impact of Stronger Capital and Liquidity Requirements”. In: *Bank of International Settlements*.
- (2011a). “Basel 3: A global regulatory framework for more resilient banks and banking systems”. In: *Bank of International Settlements*.
- (2011b). “Global systemically important banks: Assessment methodology and the additional loss absorbency requirement”. In: *Bank of International Settlements*.
- (2013). “Capital requirements for banks’ equity investments in funds”. In: *Bank of International Settlements*.
- Beltratti, A. and R. M. Stultz (2012). “The credit crisis around the globe: Why did some banks perform better?” In: *Journal of Financial Economics* 105.Issue 1, pp. 1–17.
- Benston, G. J. (1994). “Universal Banking”. In: *Journal of Economic Perspectives* 8.3, pp. 121–144.
- Benston, G. J., W. C. Hunter, and L. D. Wall (1995). “Motivations for bank mergers and acquisitions: Enhancing the deposit insurance put option versus earnings diversification”. In: *Journal of money, credit and banking* 27.3, pp. 777–788.
- Bentolila, S., P. Cahuc, J. J. Dolado, and T. Le Barbanchon (2012). “Two-tier labour markets in the Great Recession: France versus Spain”. In: *The Economic Journal* 122.562, F155–F187.
- Berger, A. N. and C. H. S. Bouwman (2013). “How does capital affect bank performance during financial crises?” In: *Journal of Financial Economics* 109.1, pp. 146–176.
- Betz, F., S. Oprică, T. A. Peltonen, and P. Sarlin (2014). “Predicting distress in European banks”. In: *Journal of Banking & Finance* 45, pp. 225–241.
- Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). “Fast unfolding of communities in large networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008+. arXiv: [0803.0476](https://arxiv.org/abs/0803.0476).
- Bongini, P., L. Nieri, and M. Pelagatti (2015). “The importance of being systemically important financial institutions”. In: *Journal of Banking & Finance* 50, pp. 562–574.
- Brunnermeier, M. K. and L. H. Pedersen (2009). “Market liquidity and funding liquidity”. In: *Review of Financial Studies* 22.6, pp. 2201–2238.
- Caliński, T. and J. Harabasz (1974). “A dendrite method for cluster analysis”. In: *Communications in Statistics-theory and Methods* 3.1, pp. 1–27.
- Chakraborty, T., S. Srinivasan, N. Ganguly, S. Bhowmick, and A. Mukherjee (2013). “Constant communities in complex networks”. In: *Scientific reports* 3.
- Chava, S., D. Livdan, and A. Purnanandam (2009). “Do shareholder rights affect the cost of bank loans?” In: *Review of Financial Studies* 22.8, pp. 2973–3004.
- Chiorazzo, V., C. Milani, and F. Salvini (2008). “Income diversification and bank performance: Evidence from Italian banks”. In: *Journal of Financial Services Research* 33.3, pp. 181–203.
- Cole, R. A. and L. J. White (2012). “Déjà vu all over again: The causes of US commercial bank failures this time around”. In: *Journal of Financial Services Research* 42.1-2, pp. 5–29.
- Damar, H. E., C. A. Meh, and Y. Terajima (2013). “Leverage, balance-sheet size and wholesale funding”. In: *Journal of Financial Intermediation* 22.4, pp. 639–662.
- Dell’Arriccia, G. and R. Marquez (2010). “Risk and the Corporate Structure of Banks”. In: *The Journal of Finance* 65.3, pp. 1075–1096.

- Demirgüç-Kunt, A. and E. Detragiache (2005). “Cross-country empirical studies of systemic bank distress: a survey”. In: *National Institute Economic Review* 192.1, pp. 68–83.
- Demirgüç-Kunt, A. and H. P. Huizinga (2010). “Bank Activity and Funding Strategies: The Impact on Risk and Return”. In: *Journal of Financial Economics* 98.Issue 3, pp. 626–650.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.
- Demyanyk, Y. and I. Hasan (2010). “Financial crises and bank failures: A review of prediction methods”. In: *Omega* 38.5, pp. 315–324.
- DeYoung, R. and K. Y. Jang (2016). “Do banks actively manage their liquidity?” In: *Journal of Banking and Finance* 66, pp. 143–161.
- DeYoung, R. and G. Torna (2013). “Nontraditional banking activities and bank failures during the financial crisis”. In: *Journal of Financial Intermediation* 22.3, pp. 397–421.
- Diamond, D. W. and R. G. Rajan (2009). “The credit crisis: conjectures about causes and remedies”. In: *American Economic Review* 99, pp. 606–610.
- Dongen, S. van and A. J. Enright (2012). “Metric distances derived from cosine similarity and Pearson and Spearman correlations”. In: *CoRR* abs/1208.3145.
- Duprey, T. and M. Lé (2014). “Bankscope dataset: getting started”. In: *Available at SSRN 2191449*.
- EBA (2014). “Guidelines: on common procedures and methodologies for the supervisory review and evaluation process (SREP)”. In: *European Banking Authority*.
- Estrella, A., S. Park, and S. Peristiani (2000). “Capital ratios as predictors of bank failure”. In: *Economic policy review* 6.2.
- Farhi, E. and J. Tirole (2012). “Collective moral hazard, maturity mismatch, and systemic bailouts”. In: *American Economic Review* 102.1, pp. 60–93. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- FASB (2016). “Comparability in International Accounting Standards: A Brief History”. In: *Financial Accounting Standards Board*. Retrieved on March 8th.
- Firth, D. (1993). “Bias reduction of maximum likelihood estimates”. In: *Biometrika* 80.1, pp. 27–38.
- Flannery, M. J. (1998). “Using market information in prudential bank supervision: A review of the US empirical evidence”. In: *Journal of Money, Credit and Banking*, pp. 273–305.
- Fortunato, S. (2010). “Community detection in graphs”. In: *Physics Reports* 486.3-5, pp. 75–174.
- Georg, C.-P. (2013). “The effect of the interbank network structure on contagion and common shocks”. In: *Journal of Banking & Finance* 37.7, pp. 2216–2228.
- González-Hermosillo, B. (1999). *Determinants of ex-ante banking system distress: A macro-micro empirical exploration of some recent episodes*. International Monetary Fund.
- Gorton, G. B. (2010). *Slapped by the invisible hand: The panic of 2007*. Oxford University Press.
- Gurvich, E. and I. Prilepskiy (2015). “The impact of financial sanctions on the Russian economy”. In: *Russian Journal of Economics* 1.4, pp. 359–385.
- Halkidi, M., Y. Batistakis, and M. Vazirgiannis (2001). “On clustering validation techniques”. In: *Journal of intelligent information systems* 17.2, pp. 107–145.
- Halling, M., J. Yu, and J. Zechner (2016). “Leverage dynamics over the business cycle”. In: *Journal of Financial Economics* 122.1, pp. 21–41.
- Han, J., M. Kamber, and J. Pei (2011). *Data mining: concepts and techniques*. Elsevier.
- Honaker, J. and G. King (2010). “What to do about missing values in time-series cross-section data”. In: *American Journal of Political Science* 54.2, pp. 561–581.
- Iannotta, G., G. Nocera, and A. Sironi (2007). “Ownership structure, risk and performance in the European banking industry”. In: *Journal of Banking and Finance* 31.7, pp. 2127–2149.
- Ivashina, V. and D. Scharfstein (2010). “Bank lending during the financial crisis of 2008”. In: *Journal of Financial Economics* 97.3, pp. 319–338. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Jeffries, I. (2011). *Economic developments in contemporary Russia*. Taylor & Francis.
- Jin, J. Y., K. Kanagaretnam, and G. J. Lobo (2011). “Ability of accounting and audit quality variables to predict bank failure during the financial crisis”. In: *Journal of Banking & Finance* 35.11, pp. 2811–2819.
- John, K., L. Litov, and B. Yeung (2008). “Corporate governance and risk-taking”. In: *The Journal of Finance* 63.4, pp. 1679–1728.

- Kashyap, A. K., R. Rajan, and J. C. Stein (2002). “Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking”. In: *The Journal of Finance* 57.1, pp. 33–73. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Khan, M. S., H. Scheule, and E. Wu (2016). “Funding liquidity and bank risk taking”. In: *Journal of Banking & Finance*, pp. 1–14.
- Kick, T. and M. Koetter (2007). “Slippery slopes of stress: ordered failure events in German banking”. In: *Journal of Financial Stability* 3.2, pp. 132–148.
- Kishan, R. P. and T. P. Opiela (2000). “Bank size, bank capital, and the bank lending channel”. In: *Journal of Money, Credit and Banking*, pp. 121–141.
- Köhler, M. (2015). “Which banks are more risky? The impact of business models on bank stability”. In: *Journal of Financial Stability* 16, pp. 195–212.
- Krause, A. and S. Giansante (2012). “Interbank lending and the spread of bank failures: A network model of systemic risk”. In: *Journal of Economic Behavior and Organization* 83, pp. 583–608.
- Krishnamurthy, A. (2010). “Amplification mechanisms in liquidity crises”. In: *American Economic Journal: Macroeconomics* 2.3, pp. 1–30.
- Laeven, L. and F. Valencia (2010). *Resolution of banking crises: The good, the bad, and the ugly*. 10-146. International Monetary Fund.
- (2012). “Systemic banking crises database: An update”. In:
- Laeven, L. and R. Levine (2009). “Bank governance, regulation and risk taking”. In: *Journal of Financial Economics* 93.2, pp. 259–275.
- Lancichinetti, A., F. Radicchi, J. J. Ramasco, and S. Fortunato (2011). “Finding statistically significant communities in networks”. In: *PloS One* 6.4, e18961.
- Lozano-Vivas, A. and F. Pasiouras (2010). “The impact of non-traditional activities on the estimation of bank efficiency: international evidence”. In: *Journal of Banking & Finance* 34.7, pp. 1436–1449.
- Manzaneque, M., E. Merino, and A. M. Priego (2015). “The role of institutional shareholders as owners and directors and the financial distress likelihood. Evidence from a concentrated ownership context”. In: *European Management Journal* 34.4, pp. 439–451.
- Mergaerts, F. and R. Vander Vennet (2016). “Business models and bank performance: A long-term perspective”. In: *Journal of Financial Stability* 22, pp. 57–75.
- Milbourn, T. T., A. W. Boot, and A. V. Thakor (1999). “Megamergers and expanded scope: Theories of bank size and activity diversity”. In: *Journal of Banking & Finance* 23.2, pp. 195–214.
- Mitchell, M. and T. Pulvino (2012). “Arbitrage crashes and the speed of capital”. In: *Journal of Financial Economics* 104.3, pp. 469–490.
- Mori, N. and N. Ikeda (2015). “Majority support of shareholders, monitoring incentive, and dividend policy”. In: *Journal of Corporate Finance* 30, pp. 1–10.
- Newman, M. E. J. and M. Girvan (2004). “Finding and evaluating community structure in networks”. In: *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* 69.2 Pt 2, p. 16.
- Oordt, M. R. C. van (2013). “Securitization and the dark side of diversification”. In: *Journal of Financial Intermediation* 23.2, pp. 214–231.
- Ostry, J. D., A. R. Ghosh, M. Chamon, and M. S. Qureshi (2012). “Tools for managing financial-stability risks from capital inflows”. In: *Journal of International Economics* 88.2, pp. 407–421.
- Paligorova, T. and J. A. C. Santos (2016). “Banks’ Exposure to Rollover Risk and the Maturity of Corporate Loans”. In: *Review of Finance*, rfw039.
- Poghosyan, T. and M. Čihák (2009). “Distress in European banks: An analysis based on a new dataset”. In: *IMF Working Papers*, pp. 1–37.
- Puliga, M., A. Flori, G. Pappalardo, A. Chessa, and F. Pammolli (2016). “The Accounting Network: How Financial Institutions React to Systemic Crisis”. In: *PLoS One* 11.10, pp. 1–14.
- Roengpitya, R., N. Tarashev, and K. Tsatsaronis (2014). *Bank business models*. BIS.
- Shleifer, A. and R. Vishny (2011). “Fire sales in finance and macroeconomics”. In: *The Journal of Economic Perspectives* 25.1, pp. 29–48.
- Stiroh, K. J. (2004). “Diversification in Banking: Is Noninterest Income the Answer?” In: *Journal of Money, Credit, and Banking* 36.5, pp. 853–882.
- Tan, P.-N., M. Steinbach, and V. Kumar (2006). *Introduction to data mining*. Vol. 1. Pearson Addison Wesley Boston.



- Tirole, J. (2011). “Illiquidity and All Its Friends”. In: *Journal of Economic Literature* 49.2, pp. 287–325.
- Traag, V. A., G. Krings, and P. Van Dooren (2013). “Significant Scales in Community Structure”. In: *Scientific Reports* 3,
- Vazquez, F. and P. Federico (2015). “Bank funding structures and risk: Evidence from the global financial crisis”. In: *Journal of Banking & Finance* 61, pp. 1–14.
- VIP (2011). *Product Review of BvD Bankscope, October 2011*.
- Wagner, W. (2010). “Diversification at financial institutions and systemic crises”. In: *Journal of Financial Intermediation* 19.3, pp. 373–386.
- Ward, J. and H. Joe (1963). “Hierarchical grouping to optimize an objective function”. In: *Journal of the American statistical association* 58.301, pp. 236–244.

# Appendix

## A Description of the Aggregated Measures

In the study of peer groups features we employ a set of aggregated measures which synthesise balance sheet items. The selection of these measures helps the interpretation of peer groups in terms of their business models because provides a less granular representation of balance sheet dimensions which, otherwise, would have make the multiple pairwise comparisons very complex. In addition, this choice partially overcomes the issues related to the presence of missing values within the set of variables used to compute the cosine similarities. The following measures are chosen among those usually applied in literature to detect business models. These aggregates are computed on standardized balance sheet variables, i.e. the constituents of each aggregated dimension are standardized by the total assets of the respective institution.

- *Retail Loans* = Residential Mortgage Loans + Other Mortgage Loans + Other Consumer/Retail Loans
- *Corporate and Other Loans* = Corporate and Commercial Loans + Other Loans
- *Retail and Corporate Loans* = Residential Mortgage Loans + Other Mortgage Loans + Other Consumer/Retail Loans + Corporate and Commercial Loans + Other Loans
- *Total Loans* = Residential Mortgage Loans + Other Mortgage Loans + Other Consumer/Retail Loans + Corporate and Commercial Loans + Other Loans + Loans and Advances to Banks
- *Interbank Lending* = Loans and Advances to Banks + Reverse Repos and Cash Collateral
- *Investments* = At Equity Investments in Associates + Available for Sale Securities + Trading Securities and At FV Through Income + Held to Maturity Securities + Other Securities
- *Customer Deposits* = Customer Deposits (Current, Savings, Term)
- *Interbank Borrowing* = Deposits from Banks + Other Deposits and Short-Term Borrowings + Repos and Cash Collateral
- *Long-Term Funding* = Senior Debt Maturing After 1 Year + Subordinated Borrowing + Other Funding
- *Long-Term Funding + Equity* = Senior Debt Maturing After 1 Year + Subordinated Borrowing + Other Funding + Total Equity
- *Wholesale Debt* = Senior Debt Maturing After 1 Year + Subordinated Borrowing + Other Funding + Other Deposits and Short-Term Borrowings + Deposits from Banks
- *Stable Funding* = Senior Debt Maturing After 1 Year + Subordinated Borrowing + Other Funding + Other Liabilities + Customer Deposits (Current, Savings, Term)
- *Stable Funding - CORE* = Senior Debt Maturing After 1 Year + Subordinated Borrowing + Customer Deposits (Current, Savings, Term)
- *Net Liquidity* = Cash and Due From Banks + Reverse Repos and Cash Collateral - Deposits from Banks - Other Deposits and Short-Term Borrowings - Repos and Cash Collateral - Customer Deposits (Current, Savings, Term)

## B Clustering Validation Comparisons

This Appendix presents some tests we run to validate our clustering approach. Clustering validation is a very complex task with no easy solutions (Han et al., 2011). We choose three main measures for clustering validation (see Subsection 5.1), such as the average silhouette width, the Pearson gamma and the average within/between ratio of the distances (see e.g. Halkidi et al., 2001). We evaluate the *Louvain* approach used in our study (based on the cosine similarity of institutions' vectors of attributes and a modularity optimization to select the best clustering configuration) against the two other reference methods used in literature, the Ward approach (which uses the euclidean distance between institutions' characteristics to assess similarities and applies the Pseudo F-Index to select the best clustering configuration) and the institutions direct classification, or specialization, as reported in the main tables.

Note that to clusterize the whole data set of institutions, Ward algorithm requires a complete set of values for institutions' characteristics to compute euclidean distances. This means that assumptions aimed at filling missing values, which are of a substantial proportion on a sample size and geographic coverage of this magnitude, have to be considered. Different assumptions on missing values would affect the clustering outcomes, making this approach not very suitable for large, complex and sparse samples. On the contrary, the cosine similarity we use in our classification, i.e. the cosine of the angle of the two vectors of institutions' characteristics, threats missing values as unknown variables and considers only the intersection of known value attributes to determine the similarities (see eq. 1). This is a major advantage compared to other approaches, such as the Ward algorithm based on euclidean distances, and motivates our choice to use the cosine similarity for this sparse and multidimensional setting (Tan et al., 2006). Finally, we decided to rely on a clustering algorithm, the *Louvain* method, that among hierarchical methodologies like the popular Ward approach, has shown to be very appropriate and efficient for complex and sparse data samples (Chakraborty et al., 2013; Lancichinetti et al., 2011).

**Table 13: Validation of Clustering Approaches.** In the table we provide clustering validation measures for the *Louvain*, Ward and direct (Bankscope) clustering methods. *avg silwidth* stands for the average silhouette widths; *pearson gamma* is the correlation between distances and a 0-1 vector where 0 means same cluster, 1 means different cluster (see Normalized gamma in Halkidi et al., 2001); *wb ratio* is the ratio of average within and average between distances. Filling of missing values: [Z] if are replaced by zeros, [A] if mean values are used, [M] if medians are applied, [EM] if a bootstrap procedure for multiple imputations based on the Expected Maximization algorithm is used (reported estimates correspond to average values over 10 simulations). Estimates are averaged over the interval 2005-14. Source: Authors' own elaborations.

	Louvain	Ward	Direct	Louvain	Ward	Direct	Louvain	Ward	Direct	Louvain	Ward	Direct
	[Z]	[Z]	[Z]	[A]	[A]	[A]	[M]	[M]	[M]	[EM]	[EM]	[EM]
avg silwidth	0.1695	0.1513	-0.0490	0.0982	0.0991	-0.0525	0.1126	0.1044	-0.0566	0.0574	0.0997	-0.0587
pearson gamma	0.1965	0.1406	0.0744	0.1272	0.1070	0.0555	0.1376	0.1107	0.0568	0.1134	0.1193	0.0686
wb ratio	0.7593	0.8050	0.9118	0.8237	0.8358	0.9313	0.8115	0.8200	0.9293	0.8555	0.8441	0.9222

Table 13 provides performance measures of each of the three clustering methodologies under four main assumptions for filling missing values (see Section 5.1): [Z] that assigns zeros to all missing values; [A] that replaces each missing value with the average value across the whole sample for that variable; [M] similar to the average case but using the medians; [EM] that estimates missing values using a bootstrap procedure for multiple imputations based on the Expected Maximization algorithm.

Looking at the values in each configuration across measures - a higher metric of the former two statistics and a smaller within-between cluster ratio provide a better fitting according to the information set - it is quite evident that direct classification shows very poor scores in clustering institutions compared to the two indirect approaches. This is consistent across all data configurations

and validation measures. Results indicate that institutions with the same direct specialization<sup>40</sup> may adopt very different business models. Therefore, this classification may not provide a useful indication of the activities institutions run, in particular when considering cross-countries comparisons. On the other side, both *Louvain* and Ward scores are much higher for the first two statistics and lower for the *wb ratio* than the direct one by far, supporting the adoption of indirect classifications as better methods to identify the true banking business models.

Between the two indirect methods, we observe a very close and consistent performance across the different model configurations and validation measures. This comparison allows us to confirm that the *Louvain* method adopted in our study does a job as good as the mainstream Ward method. However, thanks to the methodological advantage on sparse and very complex data samples, we rely on the *Louvain* algorithm since it is more suitable than Ward for the task ahead. A more detailed analysis on the emergence of peer groups and their distinct features is discussed in Appendix F, where we provide evidences on differences in balance sheet dimensions across business models by means of non-parametrically tests.

---

<sup>40</sup>Bankscope provide 16 main groups/specializations: Bank Holding & Holding Companies, Clearing Institutions & Custody, Commercial Banks, Cooperative Banks, Finance Companies (Credit Card, Factoring & Leasing), Group Finance Companies, Investment & Trust Corporations, Investment Banks, Islamic Banks, Micro-Financing Institutions, Multi-Lateral Government Banks, Other Non Banking Credit Institution, Private Banking & Asset Mgt Companies, Real Estate & Mortgage Bank, Savings Bank and Securities Firm.

## C Peer Groups Characteristics

This Appendix presents direct classification together with our clustering results to provide a helpful parallel to common banks descriptions.

### C.1 Peer Groups Balance-Sheets Features

**Table 14: Peer Groups Economic Features.** We report values for aggregated balance sheet variables standardized by total assets for groups Wholesale, Commercial and Saving. Column *Average* refers to the average values of the entire sample composed by the three groups. For variables definitions see Appendix A. Estimates are computed over the interval 2005-14. Last row provides summary statistics for Total Assets (in USD Billion). Source: Bankscope, Authors' own elaborations.

		Wholesale	Commercial	Saving	Average
	# observations	3490	2700	1915	2702
Retail Loans	<i>1st Q</i>	0.00	0.00	0.00	0.00
	<i>Median</i>	0.01	0.00	0.00	0.01
	<i>Mean</i>	0.15	0.02	0.03	0.08
	<i>3rd Q</i>	0.24	0.00	0.01	0.11
Corporate and Other Loans	<i>1st Q</i>	0.11	0.51	0.43	0.32
	<i>Median</i>	0.35	0.65	0.55	0.50
	<i>Mean</i>	0.36	0.63	0.53	0.49
	<i>3rd Q</i>	0.57	0.76	0.67	0.66
Retail and Corporate Loans	<i>1st Q</i>	0.32	0.53	0.46	0.42
	<i>Median</i>	0.58	0.66	0.59	0.61
	<i>Mean</i>	0.52	0.64	0.56	0.57
	<i>3rd Q</i>	0.73	0.77	0.69	0.73
Total Loans	<i>1st Q</i>	0.57	0.63	0.62	0.60
	<i>Median</i>	0.75	0.75	0.71	0.74
	<i>Mean</i>	0.69	0.73	0.70	0.70
	<i>3rd Q</i>	0.87	0.85	0.80	0.85
Interbank Lending	<i>1st Q</i>	0.02	0.02	0.01	0.02
	<i>Median</i>	0.11	0.06	0.09	0.09
	<i>Mean</i>	0.18	0.09	0.14	0.14
	<i>3rd Q</i>	0.25	0.12	0.21	0.20
Investments	<i>1st Q</i>	0.02	0.07	0.10	0.06
	<i>Median</i>	0.11	0.16	0.19	0.15
	<i>Mean</i>	0.17	0.19	0.21	0.18
	<i>3rd Q</i>	0.23	0.26	0.29	0.26
Customer Deposits	<i>1st Q</i>	0.06	0.02	0.72	0.21
	<i>Median</i>	0.38	0.43	0.82	0.50
	<i>Mean</i>	0.39	0.38	0.79	0.48
	<i>3rd Q</i>	0.69	0.65	0.90	0.73
Interbank Borrowing	<i>1st Q</i>	0.04	0.00	0.00	0.02
	<i>Median</i>	0.15	0.10	0.02	0.10
	<i>Mean</i>	0.23	0.17	0.06	0.17
	<i>3rd Q</i>	0.35	0.24	0.08	0.25
Long-Term Funding	<i>1st Q</i>	0.00	0.00	0.00	0.00
	<i>Median</i>	0.04	0.01	0.00	0.02
	<i>Mean</i>	0.11	0.10	0.03	0.09
	<i>3rd Q</i>	0.15	0.16	0.03	0.13
Long-Term Funding + Equity	<i>1st Q</i>	0.12	0.05	0.06	0.08
	<i>Median</i>	0.21	0.13	0.10	0.16
	<i>Mean</i>	0.29	0.19	0.12	0.22
	<i>3rd Q</i>	0.43	0.28	0.16	0.31
Wholesale Debt	<i>1st Q</i>	0.09	0.02	0.00	0.05
	<i>Median</i>	0.25	0.22	0.04	0.19
	<i>Mean</i>	0.32	0.27	0.08	0.25
	<i>3rd Q</i>	0.53	0.42	0.12	0.40
Stable Funding	<i>1st Q</i>	0.34	0.12	0.79	0.37
	<i>Median</i>	0.62	0.64	0.87	0.68
	<i>Mean</i>	0.56	0.50	0.84	0.61
	<i>3rd Q</i>	0.80	0.78	0.92	0.82
Net Liquidity	<i>1st Q</i>	-0.76	-0.80	-0.90	-0.81
	<i>Median</i>	-0.61	-0.65	-0.83	-0.67
	<i>Mean</i>	-0.54	-0.51	-0.80	-0.59
	<i>3rd Q</i>	-0.35	-0.18	-0.73	-0.38
Total Assets (USD Billion)	<i>1st Q</i>	0.13	0.18	0.51	0.24
	<i>Median</i>	0.82	0.57	1.53	0.83
	<i>Mean</i>	22.90	15.96	18.78	19.62
	<i>3rd Q</i>	4.43	2.39	6.05	4.13

**Table 15: Peer Groups Economic Features - Pre Crisis.** We report values for aggregated balance sheet variables standardized by total assets for groups Wholesale, Commercial, Saving, Long Term and Focus Retail. Column *Average (I-III)* refers to the average values within the sample composed by groups Wholesale, Commercial and Saving. Column *Average (I-V)* refers to the average values within the sample composed by all five groups. For variables definitions see Appendix A. For groups Long Term and Focus Retail, estimates refer to the interval 2005-06 (since they disappear in 2007), while for Wholesale, Commercial and Saving we consider the interval 2005-07. Last row provides summary statistics for Total Assets (in USD Billion). Source: Bankscope, Authors' own elaborations.

		Wholesale (I)	Commercial (II)	Saving (III)	Long Term (IV)	Focus Retail (V)	Average (I-III)	Average (I-V)
Retail Loans	# observations	3573	3007	1989	1475	535	2856	2116
	<i>1st Q</i>	0.00	0.00	0.00	0.00	0.62	0.00	0.03
	<i>Median</i>	0.00	0.00	0.00	0.00	0.79	0.00	0.04
	<i>Mean</i>	0.17	0.01	0.02	0.01	0.73	0.08	0.10
	<i>3rd Q</i>	0.27	0.00	0.00	0.00	0.87	0.11	0.13
Corporate and Other Loans	<i>1st Q</i>	0.10	0.51	0.43	0.45	0.01	0.32	0.32
	<i>Median</i>	0.33	0.62	0.56	0.64	0.03	0.49	0.48
	<i>Mean</i>	0.36	0.61	0.54	0.59	0.07	0.49	0.48
	<i>3rd Q</i>	0.58	0.72	0.68	0.76	0.10	0.65	0.64
Retail and Corporate Loans	<i>1st Q</i>	0.31	0.53	0.46	0.46	0.74	0.42	0.44
	<i>Median</i>	0.58	0.63	0.58	0.64	0.83	0.60	0.62
	<i>Mean</i>	0.52	0.62	0.56	0.60	0.80	0.57	0.58
	<i>3rd Q</i>	0.74	0.73	0.70	0.76	0.90	0.72	0.74
Total Loans	<i>1st Q</i>	0.62	0.64	0.63	0.66	0.81	0.63	0.64
	<i>Median</i>	0.78	0.74	0.73	0.78	0.92	0.75	0.77
	<i>Mean</i>	0.72	0.72	0.71	0.75	0.87	0.72	0.73
	<i>3rd Q</i>	0.89	0.82	0.82	0.87	0.97	0.85	0.86
Interbank Lending	<i>1st Q</i>	0.03	0.03	0.03	0.04	0.03	0.03	0.03
	<i>Median</i>	0.12	0.08	0.11	0.09	0.06	0.10	0.10
	<i>Mean</i>	0.19	0.10	0.15	0.15	0.07	0.15	0.15
	<i>3rd Q</i>	0.28	0.14	0.22	0.20	0.10	0.22	0.21
Investments	<i>1st Q</i>	0.02	0.12	0.10	0.06	0.00	0.07	0.07
	<i>Median</i>	0.11	0.20	0.19	0.14	0.03	0.16	0.15
	<i>Mean</i>	0.17	0.22	0.21	0.17	0.09	0.20	0.19
	<i>3rd Q</i>	0.24	0.29	0.29	0.24	0.13	0.27	0.26
Customer Deposits	<i>1st Q</i>	0.08	0.07	0.70	0.46	0.25	0.22	0.26
	<i>Median</i>	0.39	0.63	0.82	0.58	0.61	0.58	0.58
	<i>Mean</i>	0.40	0.48	0.79	0.57	0.49	0.52	0.52
	<i>3rd Q</i>	0.69	0.75	0.91	0.76	0.71	0.76	0.76
Interbank Borrowing	<i>1st Q</i>	0.05	0.01	0.00	0.00	0.02	0.02	0.02
	<i>Median</i>	0.16	0.11	0.03	0.03	0.06	0.11	0.10
	<i>Mean</i>	0.25	0.16	0.06	0.08	0.10	0.17	0.15
	<i>3rd Q</i>	0.39	0.20	0.09	0.10	0.15	0.25	0.23
Long-Term Funding	<i>1st Q</i>	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	<i>Median</i>	0.04	0.01	0.00	0.13	0.09	0.02	0.04
	<i>Mean</i>	0.10	0.06	0.03	0.18	0.12	0.07	0.09
	<i>3rd Q</i>	0.14	0.08	0.04	0.32	0.17	0.10	0.13
Long-Term Funding + Equity	<i>1st Q</i>	0.11	0.05	0.06	0.12	0.04	0.08	0.08
	<i>Median</i>	0.20	0.09	0.10	0.26	0.12	0.14	0.16
	<i>Mean</i>	0.25	0.13	0.13	0.30	0.16	0.18	0.20
	<i>3rd Q</i>	0.36	0.17	0.16	0.44	0.22	0.25	0.27
Wholesale Debt	<i>1st Q</i>	0.11	0.04	0.00	0.07	0.06	0.06	0.06
	<i>Median</i>	0.28	0.17	0.06	0.24	0.21	0.19	0.20
	<i>Mean</i>	0.34	0.22	0.09	0.26	0.22	0.24	0.24
	<i>3rd Q</i>	0.55	0.30	0.15	0.38	0.31	0.37	0.37
Stable Funding	<i>1st Q</i>	0.32	0.21	0.79	0.74	0.36	0.39	0.44
	<i>Median</i>	0.62	0.73	0.87	0.83	0.79	0.71	0.74
	<i>Mean</i>	0.57	0.56	0.84	0.78	0.62	0.63	0.65
	<i>3rd Q</i>	0.81	0.82	0.92	0.88	0.86	0.84	0.84
Net Liquidity	<i>1st Q</i>	-0.81	-0.87	-0.91	-0.79	-0.83	-0.85	-0.84
	<i>Median</i>	-0.67	-0.80	-0.84	-0.62	-0.73	-0.76	-0.74
	<i>Mean</i>	-0.59	-0.61	-0.81	-0.61	-0.57	-0.65	-0.64
	<i>3rd Q</i>	-0.45	-0.37	-0.75	-0.49	-0.30	-0.49	-0.48
Total Assets (USD Billion)	<i>1st Q</i>	0.11	0.15	0.40	0.18	0.19	0.19	0.19
	<i>Median</i>	0.54	0.46	1.20	0.56	0.35	0.67	0.64
	<i>Mean</i>	20.77	9.00	13.37	16.91	12.00	14.92	15.05
	<i>3rd Q</i>	3.09	1.60	4.75	2.75	0.95	2.95	2.83



**Table 16: Peer Groups Economic Features - Post Crisis.** We report values for aggregated balance sheet variables standardized by total assets for groups Wholesale, Commercial, Saving, Diversified Retail and Investments. Column *Average (I-III)* refers to the average values within the sample composed by groups Wholesale, Commercial and Saving. Column *Average (I-IV)* refers to the average values within the sample composed by the first four groups. For variables definitions see Appendix A. Estimates are computed over the interval 2008-14 (excepted for group Investments which appears only in year 2012 and 2014). Last row provides summary statistics for Total Assets (in USD Billion). Source: Bankscope, Authors' own elaborations.

		Wholesale (I)	Commercial (II)	Saving (III)	Diversified Retail (IV)	Investments	Average (I-III)	Average (I-IV)
	# observations	3454	2569	1882	2183	268	2635	2522
Retail Loans	<i>1st Q</i>	0.00	0.00	0.00	0.16	0.00	0.00	0.04
	<i>Median</i>	0.02	0.00	0.00	0.30	0.00	0.01	0.07
	<i>Mean</i>	0.15	0.02	0.03	0.34	0.01	0.08	0.14
	<i>3rd Q</i>	0.23	0.00	0.02	0.45	0.00	0.11	0.18
Corporate and Other Loans	<i>1st Q</i>	0.12	0.51	0.42	0.16	0.00	0.32	0.28
	<i>Median</i>	0.36	0.66	0.55	0.27	0.01	0.50	0.45
	<i>Mean</i>	0.37	0.63	0.53	0.27	0.05	0.49	0.44
	<i>3rd Q</i>	0.57	0.78	0.66	0.37	0.09	0.66	0.60
Retail and Corporate Loans	<i>1st Q</i>	0.32	0.53	0.46	0.50	0.00	0.42	0.44
	<i>Median</i>	0.58	0.67	0.59	0.61	0.01	0.61	0.61
	<i>Mean</i>	0.52	0.65	0.57	0.61	0.07	0.57	0.58
	<i>3rd Q</i>	0.72	0.79	0.69	0.73	0.11	0.74	0.73
Total Loans	<i>1st Q</i>	0.56	0.63	0.61	0.63	0.00	0.59	0.60
	<i>Median</i>	0.74	0.76	0.71	0.73	0.05	0.74	0.73
	<i>Mean</i>	0.68	0.73	0.69	0.72	0.10	0.70	0.70
	<i>3rd Q</i>	0.86	0.87	0.80	0.83	0.17	0.85	0.84
Interbank Lending	<i>1st Q</i>	0.02	0.01	0.01	0.05	0.01	0.02	0.02
	<i>Median</i>	0.10	0.05	0.08	0.09	0.13	0.08	0.08
	<i>Mean</i>	0.17	0.09	0.13	0.11	0.20	0.13	0.13
	<i>3rd Q</i>	0.24	0.12	0.21	0.15	0.33	0.19	0.18
Investments	<i>1st Q</i>	0.02	0.05	0.10	0.12	0.20	0.05	0.07
	<i>Median</i>	0.11	0.14	0.20	0.23	0.40	0.14	0.16
	<i>Mean</i>	0.16	0.17	0.21	0.23	0.43	0.18	0.19
	<i>3rd Q</i>	0.23	0.25	0.29	0.32	0.63	0.25	0.27
Customer Deposits	<i>1st Q</i>	0.06	0.00	0.73	0.67	0.00	0.20	0.30
	<i>Median</i>	0.38	0.35	0.82	0.75	0.00	0.47	0.53
	<i>Mean</i>	0.39	0.34	0.80	0.71	0.06	0.47	0.52
	<i>3rd Q</i>	0.69	0.60	0.90	0.81	0.08	0.71	0.73
Interbank Borrowing	<i>1st Q</i>	0.04	0.00	0.00	0.06	0.19	0.02	0.03
	<i>Median</i>	0.15	0.09	0.02	0.12	0.40	0.10	0.10
	<i>Mean</i>	0.22	0.18	0.06	0.13	0.41	0.17	0.16
	<i>3rd Q</i>	0.33	0.26	0.08	0.18	0.60	0.25	0.23
Long-Term Funding	<i>1st Q</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	<i>Median</i>	0.04	0.02	0.00	0.01	0.02	0.02	0.02
	<i>Mean</i>	0.12	0.12	0.03	0.04	0.05	0.10	0.08
	<i>3rd Q</i>	0.16	0.19	0.03	0.05	0.05	0.14	0.12
Long-Term Funding + Equity	<i>1st Q</i>	0.12	0.04	0.06	0.07	0.07	0.08	0.08
	<i>Median</i>	0.22	0.15	0.10	0.09	0.13	0.17	0.15
	<i>Mean</i>	0.31	0.22	0.12	0.12	0.17	0.23	0.21
	<i>3rd Q</i>	0.45	0.33	0.16	0.13	0.24	0.34	0.30
Wholesale Debt	<i>1st Q</i>	0.08	0.01	0.00	0.09	0.06	0.04	0.05
	<i>Median</i>	0.24	0.24	0.04	0.16	0.15	0.19	0.18
	<i>Mean</i>	0.32	0.29	0.07	0.18	0.19	0.25	0.23
	<i>3rd Q</i>	0.52	0.47	0.11	0.23	0.27	0.41	0.37
Stable Funding	<i>1st Q</i>	0.34	0.08	0.79	0.72	0.07	0.36	0.44
	<i>Median</i>	0.62	0.60	0.86	0.79	0.20	0.67	0.70
	<i>Mean</i>	0.56	0.48	0.84	0.77	0.26	0.60	0.64
	<i>3rd Q</i>	0.79	0.77	0.92	0.84	0.41	0.82	0.82
Net Liquidity	<i>1st Q</i>	-0.74	-0.77	-0.90	-0.89	-0.37	-0.79	-0.81
	<i>Median</i>	-0.58	-0.58	-0.82	-0.87	-0.22	-0.64	-0.69
	<i>Mean</i>	-0.51	-0.46	-0.79	-0.82	-0.24	-0.56	-0.62
	<i>3rd Q</i>	-0.31	-0.10	-0.73	-0.82	-0.05	-0.34	-0.44
Total Assets (USD Billion)	<i>1st Q</i>	0.14	0.19	0.55	0.31	1.97	0.25	0.27
	<i>Median</i>	0.70	0.61	1.68	0.74	7.80	0.90	0.87
	<i>Mean</i>	23.81	18.95	21.10	13.52	109.25	21.59	19.84
	<i>3rd Q</i>	5.00	2.73	6.60	2.16	52.03	4.64	4.10

## C.2 Wholesale Model

The *Wholesale* business model is the most popular model in our sample, accounting on average 3,490 institutions in the period 2005-14. The geographic map in Appendix D shows a decline in popularity among the main developed countries after the financial crisis (same result found in Roengpitya et al., 2014) compared to the opposite trend observed in developing and emerging countries. This model is characterised by the largest proportion of wholesale funding (see Table 14), from which it inherits the name. Wholesale funding accounts for 32% of total assets on average for the entire reference period. The rest of funding comes from a modest amount of deposits (39%) and relevant interbank borrowings (23%), which combined with 18% of interbank lending shows an important contribution to interbank markets as net borrower. As a result, moderate long-term funding is observed (11%). The assets side is dominated by loan investments, well spread among retail loans (15%) and corporate/commercial ones (36%). We note a moderate exposure to noninterest income investments (17%).

This model reaches the peak of popularity in 2007 with 4,108 institutions, as displayed in Table 17. Based on the number of institutions, Russia is the most represented country, representing on average 24% of the institutions which are composed by almost all commercial Russian banks, followed by US institutions (10%) which are mainly bank holdings and few investment banks, some of those forming the *Investment* peer group model after the crisis (see Section C.8). Swiss institutions used to have a marginal contribution to this group in 2005 and 2006, accounting for 5% and 3% of total numbers, respectively. Those were mainly commercial banks and private banks & asset management companies. In 2007, Swiss institutions reached 10% share in the group, due to also the migration of more than 250 Raiffeisen saving banks from another wholesale-oriented business model, the *Focus Retail* model discussed in Section C.7. Those institutions, along with small German ones, switched again in 2008 to form a new peer group running a business model named *Diversified Retail*, which will be discussed in Section C.5. Among the other main countries represented in this group, we recall the Austrian Sparkasse and Volksbanks, which are saving and commercial institutions by their direct specialization, and French banks, mainly commercial.

This business model is also the largest in terms of total assets (see Figure 4), with a total of \$61tn at the beginning of the sample and an average size of almost \$20bn per institution. For instance, we notice that in 2005 Russian institutions, although the most represented in numbers, accounted for only a small share of total assets, not even \$300bn compared with US institutions which represent almost one third of the total peer group's assets (see Appendix D). Russian commercial banks had a small average size of less than \$250mn in 2005 and represented only half of the total Russian assets in the peer group. The rest was covered by few large Russian investments, saving and real estates institutions. Opposite scenario for US bank holdings that presented average size of almost \$30bn in 2005, representing almost 70% of the whole US total assets in this group. We notice other countries not showing up by bank number representation but quite large in terms of total assets in the peer group, i.e. France (almost \$8.1tn and average size of \$33bn in 2005, mainly commercial banks by direct classification), UK (total assets of \$9.9tn in 2005 with average size equal to \$93.6bn), Netherlands (\$2.1tn and \$143.7bn on average) and Belgium (\$1.9tn and \$161.3bn average size, placing them among the largest of the peer group).

Concluding on banking direct specialization within this peer group, we note interesting geographical trends. In US the wholesale model is adopted mainly by bank holdings and few commercial banks, the latter dominant in France and Russia. On an opposite specialization category, we have the Austrian case where cooperative and saving banks adopted this business model, which is quite counter-intuitive as we would expect a more deposit-oriented business model for those banking specializations. A more volatile case for Switzerland where a large set of saving banks temporarily joined commercial and investment banks in 2007. As discussed above we just want to emphasize the issue of the direct classification approach that can lead to misleading assessments of banking activities as it does not capture the true banking business models on a global basis.

**Table 17: Wholesale.** Summary descriptive statistics over the interval 2005-14. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors’ own elaborations.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
<i>Obs.</i>	3221	3390	4108	3270	3116	3484	3608	3626	3424	3653	3490
<b>Country</b>											
RUSSIAN FEDERATION	20%	26%	22%	27%	30%	26%	24%	23%	24%	20%	24%
UNITED STATES OF AMERICA	18%	15%	13%	7%	9%	9%	10%	7%	8%	6%	10%
FRANCE	8%	6%	5%	5%	5%	5%	4%	3%	3%	3%	5%
AUSTRIA	7%	6%	6%	5%	4%	5%	3%	2%	2%	2%	4%
SWITZERLAND	5%	3%	10%	4%	1%	2%	1%	1%	1%	2%	3%
CHINA	0%	0%	1%	2%	2%	2%	3%	4%	5%	4%	2%
UKRAINE	0%	0%	0%	0%	0%	0%	1%	4%	1%	4%	1%
RoW	42%	43%	44%	50%	49%	52%	54%	56%	57%	59%	50%
<b>Specialization</b>											
Commercial Banks	51%	54%	52%	61%	62%	61%	59%	63%	61%	62%	59%
Bank Holding & Holding Companies	17%	15%	12%	8%	9%	8%	8%	8%	8%	8%	10%
Investment Banks	5%	6%	5%	6%	6%	6%	6%	5%	7%	5%	6%
Savings Bank	7%	6%	10%	6%	6%	6%	6%	5%	2%	4%	6%
Other	19%	19%	21%	20%	18%	19%	20%	19%	21%	21%	20%

### C.3 Commercial Model

The second core business model discovered by our peer group assessment, the *Commercial* one, presents a very similar liability structure to the *Wholesale* group, making it another wholesale-oriented model. Table 14 confirms the second largest wholesale funding value for this group, accounting for 27% of total assets. However, a less active participation in the interbank market, both as borrower and lender, is observed. *Commercial* model differs from the *Wholesale* group on the assets side, where the majority of the resources are invested in commercial/corporate loans (the largest among all models: 63% of total assets on average, almost twice as the *Wholesale* model) from which it inherits the name.

Table 18 confirms the dominant position of US institutions<sup>41</sup>, mainly commercial banks, with an average of 23% of the total number of institutions, quite consistent throughout the entire period. German banks used to account for the largest share of the *Commercial* model up to 2007, i.e. 59% of the total in 2005-06 and 38% in 2007. They were about two third Volksbank and Raiffeisenbank German cooperative banks and one third Sparkasse German savings banks (only about 2% of commercial and real estate institutions). This dominant participation disappeared from 2008 when many of them joined Swiss institutions in what it will be a very popular and distinctive business model, the *Diversified Retail* group, which will be discussed in Section C.5. The drop in percentages in 2007 is justified by the widening of group composition to Italian and Spanish institutions. We note that Italy has become the second largest country represented in this business model since 2007, when almost 500 Banche di Credito Cooperativo (cooperative banks) along with some commercial and few Casse di Risparmio (saving banks) switched from the *Long Term* model (discussed in Section C.6) to the *Commercial* model. They account for 23% of the total number of banks after 2007. Many Spanish cooperative and saving banks (Caixas Rural and Cajas de Ahorro) also migrated into this model in the years 2008-09.

Looking at the distribution of total assets (see Figure 4), this peer group is much smaller than the *Wholesale* one, accounting for \$14.8tn of total assets in 2005 with average size of \$5.7bn. Germany and France are the most represented countries in terms of total assets, about \$3tn in 2005 with, however, a very different individual contribution. German institutions, where saving and cooperative banks represent half of the country total assets in the peer group in 2005, are very small (\$2.3bn average size, below the group average), whereas French banks (only 45 in 2005) averaged almost \$70bn in size for that year. US institutions contribute with a smaller share of assets (\$1.5tn in 2005) with average sizes of \$2.5bn and only few very large bank holdings. Other small number of banks with large individual contributions are those from Netherlands (13 with \$131.9bn average size in 2005 that sums to an important \$1.7tn of total assets in the peer group) and UK (9 institutions with \$64.1bn average size in 2005).

<sup>41</sup>Interestingly, for example in 2005 among the 597 US banks in this peer group, 585 are US First and State banks complying with “regulatory” accounting standards.

**Table 18: Commercial.** Summary descriptive statistics over the interval 2005-14. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors’ own elaborations.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
<i>Obs</i>	2592	2568	3861	3003	2877	2760	2692	2125	2135	2390	<b>2700</b>
<b>Country</b>											
UNITED STATES OF AMERICA	23%	23%	15%	20%	24%	23%	22%	28%	27%	23%	<b>23%</b>
GERMANY	59%	59%	38%	11%	4%	4%	5%	4%	4%	4%	<b>19%</b>
ITALY	1%	2%	16%	22%	22%	23%	23%	26%	25%	22%	<b>18%</b>
FRANCE	2%	2%	2%	4%	4%	4%	5%	5%	5%	5%	<b>4%</b>
SPAIN	1%	1%	3%	5%	4%	4%	4%	2%	1%	2%	<b>3%</b>
SWEDEN	0%	0%	2%	2%	2%	2%	2%	1%	1%	3%	<b>1%</b>
AUSTRIA	0%	1%	1%	2%	2%	2%	3%	3%	3%	3%	<b>2%</b>
RoW	14%	13%	22%	35%	37%	37%	37%	32%	34%	39%	<b>30%</b>
<b>Specialization</b>											
Commercial Banks	32%	32%	32%	44%	45%	44%	43%	41%	41%	43%	<b>40%</b>
Cooperative Banks	40%	40%	39%	28%	21%	23%	23%	24%	24%	23%	<b>29%</b>
Savings Bank	18%	18%	16%	9%	8%	9%	8%	6%	7%	8%	<b>11%</b>
Finance Companies (Credit Card, Factoring & Leasing)	3%	4%	4%	6%	7%	8%	10%	13%	13%	10%	<b>8%</b>
Bank Holding & Holding Companies	1%	1%	2%	2%	4%	5%	5%	5%	5%	5%	<b>3%</b>
Securities Firm	6%	5%	7%	11%	14%	12%	11%	11%	10%	11%	<b>10%</b>

## C.4 Saving Model

The *Saving* business model is a deposit-oriented model with the most customer deposit-funded position of all, accounting for 79% of total assets as average value over the period 2005-14 (see Table 14). The remaining liability-side entries are minimal (6% interbank borrowings and 3% long-term funding) including a quite tight amount of equities (9% of total assets). On the assets side we observe a good diversification, with exposures to loans, mainly nonretail loans (of which 53% of corporate/commercial ones), noninterest income investments (21%) and decent interbank lending (14%).

The geographic composition of the *Saving* model is very stable over time with average size of 1,915 banks along the entire period, largely represented by Japanese and US institutions (30% and 27% respectively). As we mentioned above regarding the migration of Swiss (from the *Wholesale* group) and German (from the *Commercial* group) institutions to a new model (i.e. the *Diversified Retail* group), we observe a similar phenomenon here with some German banks that dropped their participation to this model in 2008. We also note the same issue with the direct bank specialization raised for the *Wholesale* classification. Focusing on those specializations, we find that among the US institutions prior to the crisis, about one third is represented by saving banks while the remaining two third is made up of commercial banks (by their direct classification). Within those commercial banks, they are almost evenly splitted among the “US GAAP” accounting standards and the “regulatory” principles. Japanese institutions, however, are mainly cooperative banks (almost 80% of the total Japanese institutions) with marginal presence of commercial banks (less than 20%), in line with the German Raiffeisenbank and Volksbank institutions (cooperative) with only marginal representation of commercial ones. Another well represented country is India, with the majority of commercial Indian banks adopting this business model. Even though there is usually consistency among direct specializations of institutions within the same country, less meaningful cross-country comparisons can be done based on the direct specialization approach.

With more than \$20tn worth of total assets (see Figure 4), *Saving* model institutions are also the second largest peer group in our global sample prior to the crisis, with average size of \$11bn in 2005. Differently from the previous two core models, the geographic representation of total assets is in line with the model popularity, having again Japanese and US banks contributing quite evenly to the total assets (\$10.3tn and \$8.5tn in Japan and US respectively in 2005) with averages within each country slightly above the peer group level (\$17.3bn and \$15.2bn in Japan and US respectively for 2005). As we would expect, saving and cooperative banks in both countries tend to be on average much smaller than the commercial ones. Other remaining countries, like Germany with total assets of \$113.1bn in 2005, contribute very marginally to the total assets of the group.

**Table 19: Saving.** Summary descriptive statistics over the interval 2005-14. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors' own elaborations.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
<i>Obs</i>	1946	1952	2070	1708	2125	2009	1778	1901	1849	1807	1915
<b>Country</b>											
JAPAN	31%	30%	28%	33%	27%	28%	31%	30%	30%	30%	30%
UNITED STATES OF AMERICA	29%	27%	26%	29%	28%	27%	24%	28%	28%	27%	27%
GERMANY	6%	7%	9%	4%	2%	2%	2%	2%	2%	2%	4%
INDIA	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%
INDONESIA	2%	2%	2%	1%	2%	2%	2%	2%	3%	3%	2%
PORTUGAL	0%	0%	0%	0%	0%	0%	1%	4%	5%	5%	1%
DOMINICAN REPUBLIC	2%	0%	0%	2%	2%	2%	2%	2%	0%	2%	1%
RoW	29%	31%	32%	28%	37%	37%	34%	29%	29%	28%	31%
<b>Specialization</b>											
Commercial Banks	50%	50%	47%	48%	49%	47%	48%	44%	42%	43%	47%
Cooperative Banks	30%	30%	30%	30%	24%	25%	28%	26%	26%	27%	28%
Savings Bank	10%	10%	10%	10%	9%	9%	6%	11%	11%	11%	10%
Bank Holding & Holding Companies	2%	3%	5%	5%	10%	11%	10%	12%	12%	11%	8%
Other	7%	7%	8%	7%	8%	8%	8%	8%	8%	8%	8%

## C.5 Diversified Retail Model

The *Diversified Retail* is a deposit-oriented model mainly formed by German and Swiss institutions which moved from the two wholesale-funded models, the *Commercial* and *Wholesale* respectively. We recall that these institutions are mainly German and Swiss Volksbanks and Raiffeisenbanks (mostly cooperative institutions) along with Sparkasse-type German saving banks. The liabilities side looks more retail-oriented than the *Wholesale* group (see Table 16), with the customer deposits size twice as big as the wholesale-funded models (71% of total assets) but not as large as the *Saving* one. Also, wholesale funding and interbank borrowings (18% and 13%) are lower than the *Wholesale* and *Commercial* models. The assets side is well diversified with a relevant exposure to retail loans (34% of total assets), corporate/commercial ones (27%), moderate interbank lending (11%) as well as decent noninterest income investments (23%).

Regarding the geographic composition of this group, Table 20 shows that US, after being the second largest popular country in 2008, felt in participation from 13% to 2%. Looking at the specialization of US institutions in 2008, among the total of 257 institutions we notice a prevalence of bank holdings (220), which present different accounting standards: 122 institutions adopted US GAAP while 98 were under regulatory principles. The former were institutions coming from the *Wholesale* group in 2007 and switching into the *Saving* model (50%) or remaining in the *Diversified Retail* group (30%) in 2009; the latter were again institutions which belonged to the *Wholesale* group in 2007 and that in 2009 migrated more to the *Commercial* model (60%) and the *Saving* model (40%). In addition, institutions under the US GAAP standards were basically within the *Wholesale* model even in the interval 2005-06, while in the same biennium those under the regulatory accounting standards were almost evenly splitted among the *Wholesale* group (40%) and the *Focus Retail* (see Section C.7) group (60%).

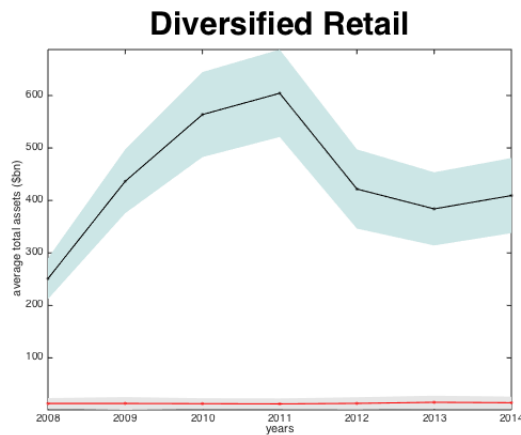
The popularity of this new peer group is also confirmed by the share of the total assets (see Figure 5) at about \$27tn in 2008 and average size of the institutions of \$13bn. In 2008, the main contributors were institutions from US with \$6.8tn worth of total assets, followed by Germany (\$4.8tn) and France (\$4.5tn). Swiss banks were relatively very small with average size of less than \$2bn in 2008 compared with the US ones with \$26.5bn. We also note that few of the US bank holdings in this group in 2008 were huge institutions with total assets greater than \$1tn. From 2009 onwards the US share in total assets lost 72% to a total of less than \$2tn in 2009 (about \$4tn from 2010 onwards) due to both a huge drop in numbers and the departure of big institutions. Among German institutions, we observe a consistent dispersion in the distribution of assets in 2008, with cooperative and saving at the lower end of the distribution (average size of cooperative at \$780mn, \$3.1bn for savings) compared with few commercial, bank holdings and some huge credit financial companies with average size of almost \$38bn. Few distressed companies adopted this business model at the peak of the crisis in 2008: 15 that mainly came from the *Wholesale* group (only 2

from *Commercial*). Those were huge compared to the average size of their peers (average size of \$251bn in 2008, almost twenty times larger than the average of the group at that time). Again we have evidence showing that relative size of their peer groups matters in terms of vulnerability and probability of distress. However, we note that those companies used to adopt a wholesale-oriented model prior to the crisis and should be compared with those peer group members in the assessment of the risk of distress (as we have shown in the sections above).

**Table 20: Diversified Retail.** Summary descriptive statistics over the interval 2008-2014. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors' own elaborations.

	2008	2009	2010	2011	2012	2013	2014	Average
<i>Obs</i>	2053	2195	2233	2252	2244	2218	2086	<b>2183</b>
<b>Country</b>								
GERMANY	62%	71%	70%	68%	69%	67%	68%	<b>68%</b>
SWITZERLAND	13%	12%	11%	11%	11%	11%	12%	<b>12%</b>
UNITED STATES OF AMERICA	13%	2%	2%	4%	2%	3%	2%	<b>4%</b>
UNITED KINGDOM	3%	2%	3%	3%	2%	2%	3%	<b>3%</b>
AUSTRIA	1%	3%	2%	3%	3%	3%	2%	<b>2%</b>
FRANCE	1%	1%	1%	1%	2%	3%	3%	<b>2%</b>
CANADA	0%	0%	1%	1%	1%	1%	1%	<b>1%</b>
RoW	9%	9%	10%	9%	9%	10%	9%	<b>9%</b>
<b>Specialization</b>								
Cooperative Banks	40%	47%	45%	44%	44%	44%	43%	<b>44%</b>
Savings Bank	30%	32%	33%	36%	35%	35%	35%	<b>34%</b>
Commercial Banks	11%	10%	11%	11%	11%	11%	11%	<b>11%</b>
Real Estate & Mortgage Bank	5%	4%	5%	4%	4%	4%	4%	<b>4%</b>
Other	14%	6%	7%	5%	5%	6%	6%	<b>7%</b>

**Figure 5: Total Assets Distribution: Diversified Retail Model.** Red curve represents the time series of average total assets (in US Billion) for all Diversified Retail institutions, while blue curve stands for distressed institutions only. Dispersion area refers to  $\pm 0.1\sigma$ . Membership is updated yearly. Source: Bankscope, Authors' own elaborations.



## C.6 Long Term Model

The *Long Term* group was a wholesale-oriented model quite popular between European banks prior to the financial crisis with an average size of 1,475 banks. It is discovered by our classification approach only in 2005 and 2006. Italian banks are the most represented by far, accounting for 43% and 40% in 2005 and 2006 respectively. Spanish banks are the second most popular with 7% and 6%. Interestingly, about 80% of those Italian and Spanish institutions are cooperative (Banche di Credito Cooperative and Cajas Rural) and saving (Casse di Risparmio and Cajas de Ahorros) institutions that migrated mainly to *Commercial* model in 2007. The remaining are commercial banks by direct specialization.



Looking at Table 15 we note very similar loans exposure with the *Commercial* model, large positions to commercial/corporate loans (59%) and noninterest income investments (17%). The liabilities side presents a large wholesale funding (26%) and modest deposits (57%) in line with the two wholesale-funded models. However, this model discriminates itself from the others by the largest long-term funding of all, 18% of total assets, that motivates the choice of the name. It is indeed the most long-term funded model among all wholesale-oriented models. Another interesting peculiarity of the model is the net lender position in the interbank market, a characteristic we observe only in the *Saving* model.

This peer group was as large as the *Commercial* model in 2005, accounting for \$19.5tn of assets (it increases to \$30.8tn in 2006). Institutions in this peer group show average size of \$13.7bn in 2005 (\$20.1bn in 2006), more than twice the size of the *Commercial* group but not as big as the *Wholesale* one (see Figure 4). The geographic concentration of assets in 2005 sees Italy with a total of almost \$4tn in 2005 due to the popularity of this model among Italian banks, followed by UK (\$4tn), France and Switzerland (\$1.6tn), Belgium (\$1.5bn) and Spain (about \$1bn). In 2006, UK institutions accounted for almost \$7tn, followed by Italy and France (\$4.5tn), Germany, Switzerland and Belgium (about \$2tn).

**Table 21: Long Term.** Summary descriptive statistics over the interval 2005-2006. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors' own elaborations.

	2005	2006	Average
<i>Obs</i>	1416	1534	1475
<b>Country</b>			
ITALY	43%	40%	41%
SPAIN	7%	6%	7%
SWEDEN	5%	5%	5%
GERMANY	2%	4%	3%
UNITED KINGDOM	3%	3%	3%
FRANCE	2%	3%	2%
SWITZERLAND	0%	2%	1%
RoW	37%	37%	37%
<b>Specialization</b>			
Cooperative Banks	38%	36%	37%
Commercial Banks	34%	34%	34%
Savings Bank	11%	10%	11%
Private Banking & Asset Mgt Companies	3%	5%	4%
Securities Firm	14%	14%	14%

## C.7 Focus Retail Model

This is another wholesale-oriented model characterized by a good proportion of wholesale funding (22% of total assets), few deposits (49%) and interbank borrowings (10%). The peculiarity of this model is the largest exposure to retail loans, from which it inherits the name, accounting for 73% of total assets. Only 7% of loans are provided to corporate/commercial instruments as well as small positions in noninterest income investments (only 9%). We also note a very small amount of equity (4% of total assets on average) that could be justified by the investment strategy (see Table 15). In fact, retail loans tend to have on average lower risk weighting than non-retail loans (especially if claims are on companies below A- credit ratings) that would justify this result<sup>42</sup>.

The geographic composition of this model is dominated by Swiss saving banks (Raiffeisenbanks) and US bank holdings under “regulatory” accounting standards in 2005. It is interesting to note how accounting standards affect the adoption of business models among the US bank holdings. Due to their geographical restriction on local investments only, those loan investments are mainly dedicated to retail customers, whereas the other bank holdings on GAAP accounting standards can better diversify their investment strategy and joining the peer group adopting the wholesale model. In 2006 some of the US institutions began their migration to the other wholesale-oriented models. We note here that a large proportion of Swiss banks moved to the *Wholesale* model in 2007, right before the restructuring process of 2008 discussed in Section C.5 where they finally

<sup>42</sup>For a deeper analysis on banks capital requirements, see reports in <http://www.bis.org/publ/bcbsca.htm>.

landed. Among the US bank holdings, most of them moved to the *Wholesale* group in 2007 and then changed to the *Diversified Retail* group in 2008 and again to the *Commercial* group in 2009. Restructuring activities among those institutions could be caused by the recovering process of the post financial crisis. A similar analysis can be done for the UK banks within this business model. Those where mainly UK real estate & mortgage institutions that moved to the *Wholesale* model in 2007 (as the closest wholesale-oriented group). However, at the peak of the financial crisis, all of them switched to the *Diversified Retail* model showing a drastic and rapid funding transformation towards deposit-oriented sources. The housing market crisis is the obvious explanation of these transitions that pushed those institutions to reduce their wholesale debt, probably due to the lack of funding opportunities.

This peer group does not exceed in assets size (see Figure 4) due to its lack of popularity compared with the other wholesale-oriented models. In 2005, the group accounted for \$5.5tn of total assets (\$6.9tn in 2006), mainly covered by institutions from the UK (\$2.2tn in 2005 and \$2.7tn in 2006) and the Netherlands (\$1tn in 2005 and \$1.2tn in 2006). The average size of the institutions in this group is \$9bn (\$14.9 in 2006), in line with the other models of this kind. However, the distribution of assets among the main represented countries varies a lot, from large UK real estate & mortgage banks (average \$8.3bn in 2005) and a couple of very big bank holdings (on average \$374bn in 2005) to very small Swiss saving banks (\$238mn on average in 2005) and US institutions (\$1.3bn on average in 2005).

**Table 22: Focus Retail.** Summary descriptive statistics over the interval 2005-2006. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors' own elaborations.

	2005	2006	Average
<i>Obs</i>	606	463	535
<b>Country</b>			
SWITZERLAND	46%	60%	53%
UNITED STATES OF AMERICA	33%	14%	23%
UNITED KINGDOM	9%	10%	9%
GERMANY	2%	3%	3%
NETHERLANDS	1%	2%	2%
SPAIN	1%	2%	1%
AUSTRALIA	0%	1%	0%
RoW	7%	8%	8%
<b>Specialization</b>			
Savings Bank	37%	48%	42%
Bank Holding & Holding Companies	34%	15%	25%
Commercial Banks	12%	16%	14%
Real Estate & Mortgage Bank	13%	15%	14%
Other	4%	6%	5%

## C.8 Investment Model

The last business model we present here is the *Investment* model, which only runs in 2012 and 2014. Table 16 shows that this model does not belong to either deposits or wholesale oriented groups. It resembles the characteristics of investment institutions with liabilities dominated by interbank borrowings (41%), the largest position ever encountered in our business classification. Along with a minimal long-term funding (only 5%) and customer deposits (6%), the investment model shows a unique no-stable and short-term funding structure. The funding is then invested mainly in noninterest income securities (43% of total assets), making it the right candidate for a trading/investment oriented model consistent with those found in Roengpitya et al. (2014)<sup>43</sup> and Ayadi et al. (2012)<sup>44</sup>. *Investment* model is also characterized by a 20% interbank lending, placing institutions in this group among the largest net borrowers in the interbank market.

<sup>43</sup>They present a trading model characterized by 19% interbank borrowing, much smaller than ours but still representing the model with the largest amount compared with the retail and wholesale ones, as well as the largest trading position.

<sup>44</sup>Their investment model is characterized by the largest position in trading assets, similar to ours. However, they do not find evidences of either dominant interbank activities or non-deposit funding. We note here that their sample does only consider large EU banks, which does not provide a good match with the geographic composition of our *Investment* model.

As we expected, the geographic representation of the *Investment* group presented in Table 23 includes institutions located in traditional financial markets, such as the US with 19% of total number of institutions, UK (12%), Taiwan (8%) and Japan (7%), representing the top broker dealer institutions. This is a very specialized peer group in size, with only 310 institutions in 2012 and 226 in 2014. All those institutions moved back and forth from and to the *Wholesale* model with only very few exceptions.

The other distinctive characteristic of this model is the size of the institutions, way larger than any other model we discover. In 2012 the 310 banks accounted for \$33.2tn of total assets with average size of \$107.4bn, at least 5 times the size of its competitor groups, with few companies with more than \$2tn of total assets in that year. The geographic map locates UK institutions whose assets sum to \$9tn, followed by US (\$7.7tn) and Japan (almost \$2tn). Many of those huge institutions are bank holdings and investment banks by their direct classification. We observe a decrease in total assets in 2014 (\$25.1tn) due to the reduction of 25% of institutions present in the group with respect to 2012. However, the average size remained very high (\$111.3tn) placing this model at the top for individual total assets. As this model emerges only well after the financial crisis, no useful information regarding the extent of risk can be retrieved. We can only observe three large institution that survived from a distress event during the crisis and then adopted this model. Two came from the *Wholesale* model and one was first in *Long Term* in 2005-06 and then in *Wholesale*. Those institutions have average assets size pretty close to \$1tn, that might suggest high vulnerability to future crisis.

**Table 23: Investments.** Summary descriptive statistics for 2012 and 2014. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period. Source: Bankscope, Authors' own elaborations.

	2012	2014	Average
<i>Obs</i>	310	226	268
<b>Country</b>			
<b>Country</b>			
UNITED STATES OF AMERICA	21%	18%	19%
UNITED KINGDOM	11%	12%	12%
MEXICO	8%	9%	8%
TAIWAN	7%	9%	8%
REPUBLIC OF KOREA	3%	11%	7%
JAPAN	6%	7%	7%
BRAZIL	6%	4%	5%
RoW	38%	29%	33%
<b>Specialization</b>			
<i>Investment Banks</i>	28%	36%	32%
<i>Commercial Banks</i>	24%	20%	22%
<i>Securities Firm</i>	23%	21%	22%
<i>Bank Holding &amp; Holding Companies</i>	11%	9%	10%
<i>Other</i>	15%	14%	14%

## C.9 Mixed Groups

We briefly comment here on the characteristics of three mixed groups adopted by a very volatile set of institutions worldwide (see Section D for details.) during the periods 2011-2013. The liabilities side shows a well mixed composition between wholesale and costumer deposits that places the funding strategy of those institutions in between a deposit-oriented and wholesale-oriented business models. The assets side is quite peculiar in 2011 with by far the largest exposure to interbank lending among all business models we find. This could be the evidence of sluggish credit markets where financial institutions were reluctant of investing into the real economy. As a result, large surplus of liquidity were injected into the interbank market. This characteristics tends to fade away towards 2013 in favour of a more traditional wholesale-oriented model.

**Table 24: Mixed groups.** We report values for aggregated balance sheet variables standardized by total assets for the Mixed groups. For variables definitions see Appendix A. Estimates are computed for each year over the interval 2011-13. Last row provides summary statistics for Total Assets (in USD Billion). Source: Bankscope, Authors' own elaborations.

		2011	2012	2013
	# observations	573	1053	1424
<b>Retail Loans</b>				
	<i>1st Q</i>	0.00	0.00	0.00
	<i>Median</i>	0.00	0.00	0.06
	<i>Mean</i>	0.05	0.05	0.26
	<i>3rd Q</i>	0.02	0.02	0.53
<b>Corporate and Other Loans</b>				
	<i>1st Q</i>	0.03	0.12	0.04
	<i>Median</i>	0.11	0.30	0.17
	<i>Mean</i>	0.13	0.32	0.22
	<i>3rd Q</i>	0.21	0.49	0.35
<b>Retail and Corporate Loans</b>				
	<i>1st Q</i>	0.06	0.17	0.26
	<i>Median</i>	0.15	0.36	0.48
	<i>Mean</i>	0.18	0.37	0.48
	<i>3rd Q</i>	0.27	0.55	0.70
<b>Total Loans</b>				
	<i>1st Q</i>	0.41	0.44	0.52
	<i>Median</i>	0.65	0.65	0.71
	<i>Mean</i>	0.61	0.61	0.66
	<i>3rd Q</i>	0.85	0.80	0.84
<b>Interbank Lending</b>				
	<i>1st Q</i>	0.21	0.06	0.02
	<i>Median</i>	0.45	0.18	0.10
	<i>Mean</i>	0.44	0.25	0.19
	<i>3rd Q</i>	0.66	0.39	0.29
<b>Investments</b>				
	<i>1st Q</i>	0.02	0.04	0.03
	<i>Median</i>	0.11	0.13	0.11
	<i>Mean</i>	0.17	0.17	0.15
	<i>3rd Q</i>	0.26	0.25	0.21
<b>Customer Deposits</b>				
	<i>1st Q</i>	0.47	0.63	0.54
	<i>Median</i>	0.71	0.76	0.71
	<i>Mean</i>	0.61	0.72	0.64
	<i>3rd Q</i>	0.83	0.84	0.82
<b>Interbank Borrowing</b>				
	<i>1st Q</i>	0.01	0.00	0.01
	<i>Median</i>	0.06	0.04	0.05
	<i>Mean</i>	0.17	0.08	0.11
	<i>3rd Q</i>	0.25	0.11	0.15
<b>Long-Term Funding</b>				
	<i>1st Q</i>	0.00	0.00	0.00
	<i>Median</i>	0.00	0.00	0.00
	<i>Mean</i>	0.02	0.01	0.06
	<i>3rd Q</i>	0.00	0.01	0.05
<b>Long-Term Funding + Equity</b>				
	<i>1st Q</i>	0.07	0.08	0.10
	<i>Median</i>	0.12	0.12	0.14
	<i>Mean</i>	0.15	0.15	0.20
	<i>3rd Q</i>	0.19	0.18	0.24
<b>Wholesale Debt</b>				
	<i>1st Q</i>	0.01	0.01	0.02
	<i>Median</i>	0.07	0.04	0.08
	<i>Mean</i>	0.19	0.09	0.16
	<i>3rd Q</i>	0.25	0.13	0.21
<b>Stable Funding</b>				
	<i>1st Q</i>	0.53	0.71	0.64
	<i>Median</i>	0.76	0.80	0.79
	<i>Mean</i>	0.66	0.77	0.73
	<i>3rd Q</i>	0.86	0.87	0.86
<b>Net Liquidity</b>				
	<i>1st Q</i>	-0.83	-0.83	-0.79
	<i>Median</i>	-0.70	-0.71	-0.67
	<i>Mean</i>	-0.63	-0.63	-0.60
	<i>3rd Q</i>	-0.50	-0.52	-0.49
<b>Total Assets (USD Billion)</b>				
	<i>1st Q</i>	0.00	0.00	0.00
	<i>Median</i>	0.23	0.27	0.27
	<i>Mean</i>	13.17	13.71	13.91
	<i>3rd Q</i>	1.56	1.71	1.80

## D Geographic Map of Business Models

This Appendix extends the discussion of the business models of Section 6 by focusing on banking activities of the main countries. Figures 6-12 presents the annual composition of peer groups, both in terms of number of institutions and total assets of each peer group, within each main country for the whole period 2005-2014. Annual peer group sizes are also compared with the total Bankscope spectrum of institutions for that country in each year (black dashed line). First thing to notice is the extremely high total assets coverage of institutions we managed for our classification study, despite the lack of consistency of data availability for some institutions/countries that were excluded from our analysis (see Footnote 9).

The US banking system is the largest system worldwide with total assets exceeding \$40tn in 2007<sup>45</sup>. US banks experienced one main falls at \$36tn right at the peak of the financial crisis in 2008 with a sluggish recovery affected by the US debt ceiling crisis in 2011 (Figure 6 on the right, dashed line). At the bank level, we note a constant reduction of number of banks, due to defaults and merges caused by the financial crisis, from more than 2000 institutions in 2005 to 1400 in 2014. Although the US banking system appears to be very well diversified and balanced in terms of business models, the fall in numbers has mainly impacted banks adopting the *Wholesale* model, the largest in terms of total assets. Those few *Focus Retails* banks initially moved to *Diversified Retail* in 2008 for then mainly split between the *Commercial* and *Saving* model right after the crisis. The fall of *Wholesale* banks has also affected the composition of the US total assets, promoting more traditional business models like the *Saving* and the *Diversified Retail* from 2008. The latter point is again evidence of relevant restructuring of banking activities towards less risky strategies as also highlighted in Roengpitya et al. (2014). We also note the emergence of the *Investment* models in 2012 and 2014 adopted by few very large banks previously adopting the *Wholesale* model.

Opposite scenario for the Russian banking system, characterized by almost all *Wholesale* banks, with few *Commercial* and *Saving* institutions emerging in the post financial crisis. Although the number of banks fluctuated among 900 institutions (see Figure 6), we note a speedy growth in total assets from \$300bn in 2005 to almost \$2tn on 2013 in Fig 6. Russian banks experienced a fast recovery from the 1998 Russian crisis thanks to the boom of oil price and the cash injection by the government into the economy to promote growth (Jeffries, 2011). In 2014 the sanction imposed by the US government had negative repercussion to the banking system contributing to a fall of 20% of banks total assets (for a broader picture, see Gurvich and Prilepskiy, 2015).

In Europe, the UK banking system reached almost \$30tn in 2008 in Figure 6, even though those assets are owned by a very small number of institutions (380 in 2005 raising up to almost 500 in 2012 as shown in Figure 6 on the left). We note that a decent share of very small UK institutions were excluded by our classification approach due to lack of data coverage across balance sheet variables (see Footnote 9). This is confirmed by the fact that the total assets coverage is extremely high (over 90%) as for all other countries. The financial crisis hit the UK banks mainly from 2009 with a loss of 10% of assets that deepened in recent years. Like the US system, UK banks tend to be quite diversified in terms of business models, with predominant wholesale-oriented funding structure. Indeed, only few banks were adopting the *Saving* model (< 10%) whereas the others adopted mainly the *Wholesale* and the *Commercial* business models. We note that almost half of the banks used to adopt *Long Term* or *Focus Retail* models in 2005-06, for then switching to mainly the *Commercial* model in 2007. Part of them (together with some *Wholesale* institutions) moved back into the hybrid *Diversified Retail* model in 2008. In terms of total assets, we note a fall in *Wholesale* models after having reached the peak in 2008 with more than \$20tn. The year after there was an evident change of activity towards *Commercial* and *Diversified Retail* models showing an important impact in assets allocation from a well diversified investment strategy (the *Wholesale*) to a more specific lending strategy promoting just commercial and retail loans. Like the US, few giant *Wholesale* institutions switched into *Investment* model in 2012 and 2014. Finally, a good proportion of institutions adopted in the period 2011-2013 the *Mixed* strategy that is briefly discussed in Appendix C.9.

The case of Swiss banks is interesting as they experienced a big transformation after the 2008

---

<sup>45</sup>Values are the representation of Bankscope database, which is stated to cover more than 90% of the total assets values (VIP, 2011).

financial crisis, moving from wholesale-oriented to deposit-oriented business models. With few exceptions, all the 450 banks in Switzerland were adopting a wholesale-oriented model, mainly *Focus Retail* banks, few big *Long Term* and the rest *Wholesale* in 2005-06. In 2007 all wholesale-oriented institutions converged together into the *Wholesale* business model to then finally landing to the *Diversified Retail* group (Figure 6 on the left). However, those banks represented a very small portion of the total assets of the Swiss banking sector, mainly saving and cooperative banks. As shown in Figure 6 (on the right), *Wholesale* model dominated the 2007-2011 period in terms of total assets. Last three years show the dominance of few huge banks adopting the *Investment* model (in 2012 and 2014) and the presence of some institutions belonging to the *Mixed* group.

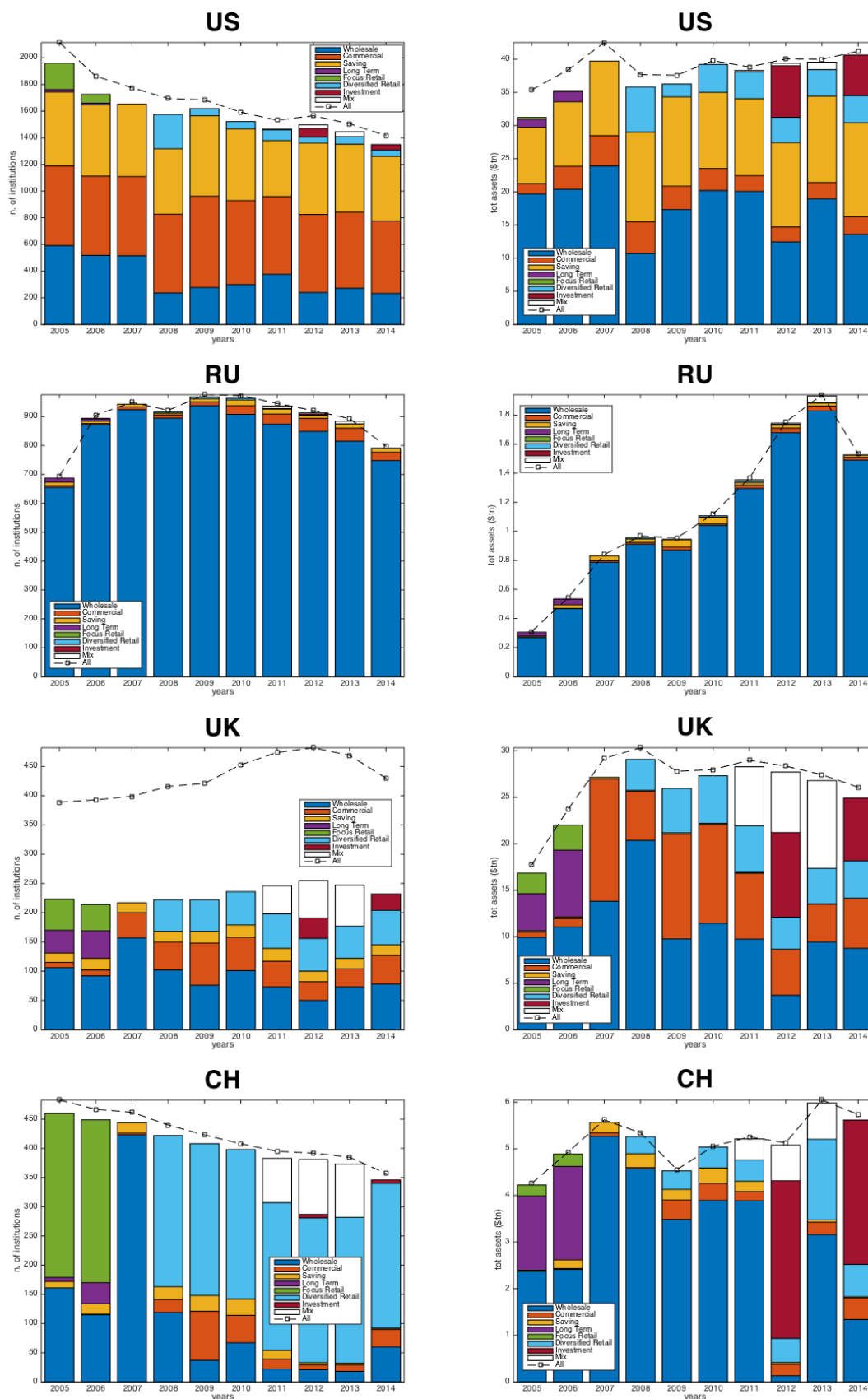
Among the main Eurozone countries plotted in Figure 7, Germany takes the lead of the largest banking sector in both number of banks (almost 2000 throughout the period with very low variation) and total assets, experiencing a fast growth by tripling its bank assets value from \$6tn in 2005 to \$18tn in 2009. As already discussed in Section 6, German banks used to adopt predominately the *Commercial* model along with small Swiss cooperative and saving banks, joined together a new peer group characterized by the *Diversified Retail* business model. This new business model became the most popular among German banks both in terms of both number of banks adopting the model and of total assets. French banks, as second largest banking system in the Eurozone, witnessed a decline in the *Wholesale* activities towards *Commercial* and, marginally, *Diversified Retail* models, both in terms of number of banks and total assets. Among the peripheral Eurozone countries, Italian and Spanish banks show a decent growth in banking assets that peaked in 2009-10 at historical levels of more than \$6tn each. This process reversed as the Eurozone faced the so-called Eurozone sovereign debt crisis starting in 2010-11 with the first Greek bailout. In both countries, the majority of banks used to adopt a *Long Term* model in 2005-06 period. From 2007 onwards, Italian and Spanish institutions converged to *Commercial* business model. A large proportion of Spanish banks, however, moved in 2009 to a more traditional and conservative banking model, the *Saving* one. This transition could be explained by the weakness of the Spanish banking system, and the whole economy, that was experiencing deterioration of banks credit worthiness, high inflation rates, unemployment and huge public debt. These events resulted in the Spanish bailout in 2012 (see e.g. Bentolila et al., 2012).

In the Asian area pictured in Figure 8, most of the emerging countries show a fast and consisted growth over the years, led by China with a total assets expansion over ten times during the last decade. Like Russia, almost all Chinese institutions (including Hong Kong) constantly adopted the *Wholesale* model. An opposite and unique business model is adopted in Japan (with peaks of almost \$30tn of total assets) and India, in which the vast majority of institutions belong to the *Saving* group.

For completeness, we also report evidence of convergence to the Asian area business model preferences in the Latin America countries (on a much smaller total assets magnitude as plotted in Figure 9) compared to the East European countries that show business model patterns more similar to the Eurozone members (more diversified adoption of business models, Figure 10). Finally, we report in Figure 11) as well as the major offshore financial centres (in Figure 12).

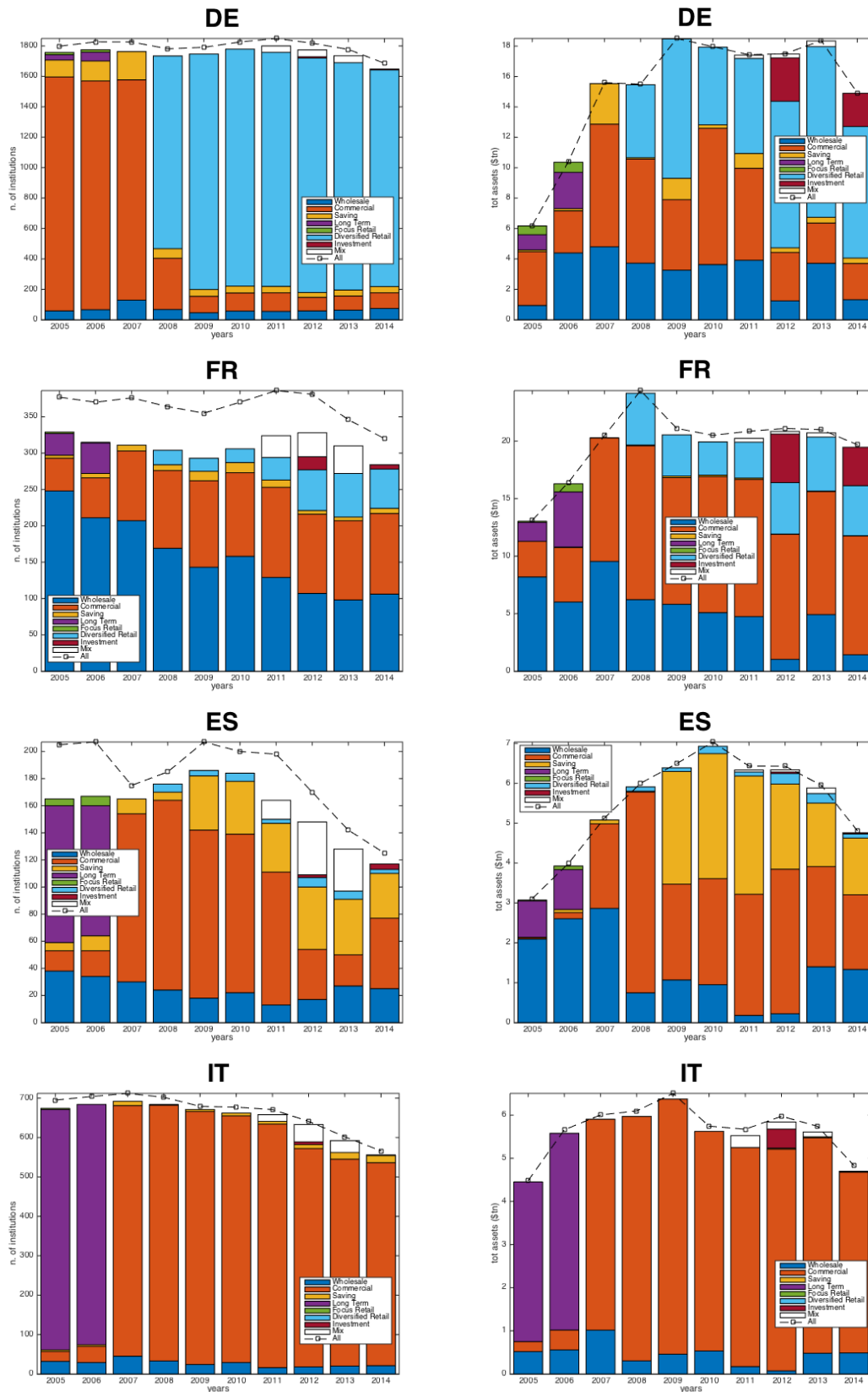
To sum up, we recall evident differences in business model preferences among countries, with the US banks being more balanced between wholesale and deposit oriented models whereas EU countries prefer a more wholesale oriented business strategy, in particular the *Commercial* business model that seems a distinctive characteristic of European institutions. The financial crisis, however, forced banks in developed countries to invest in less risky and more traditional banking activities across the globe, promoting more deposit oriented business models instead of wholesale oriented. Opposite scenario in China, Russia and Latin American countries in which the speedy growth of banking assets was dominated by institutions adopting wholesale-oriented models.

**Figure 6: Main Countries Business Models.** The plot shows the composition of banking business models in the United States (US), Russian Federation (RU), the United Kingdom (UK) and Switzerland (CH) in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Source: Bankscope, Authors' own elaborations.

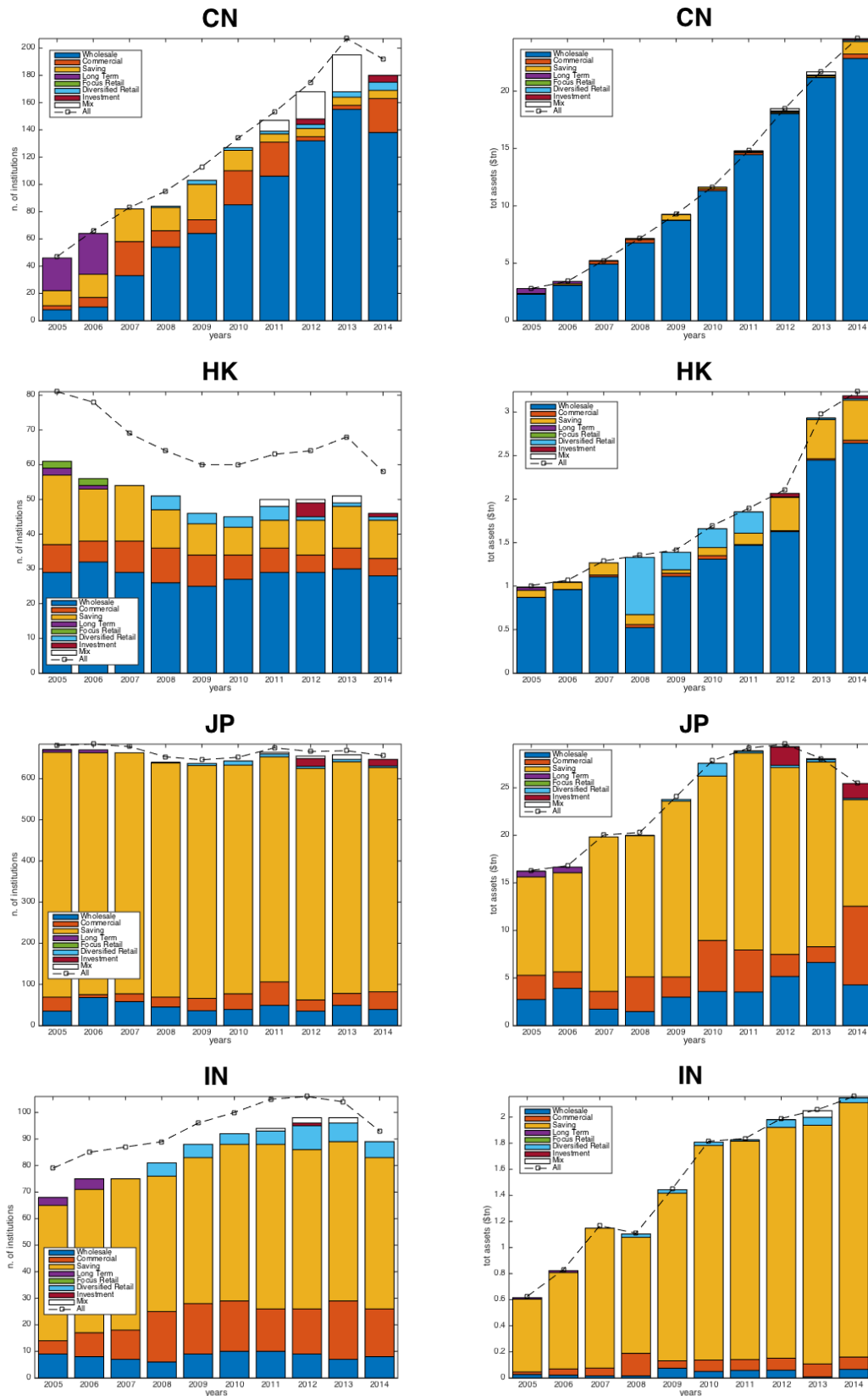




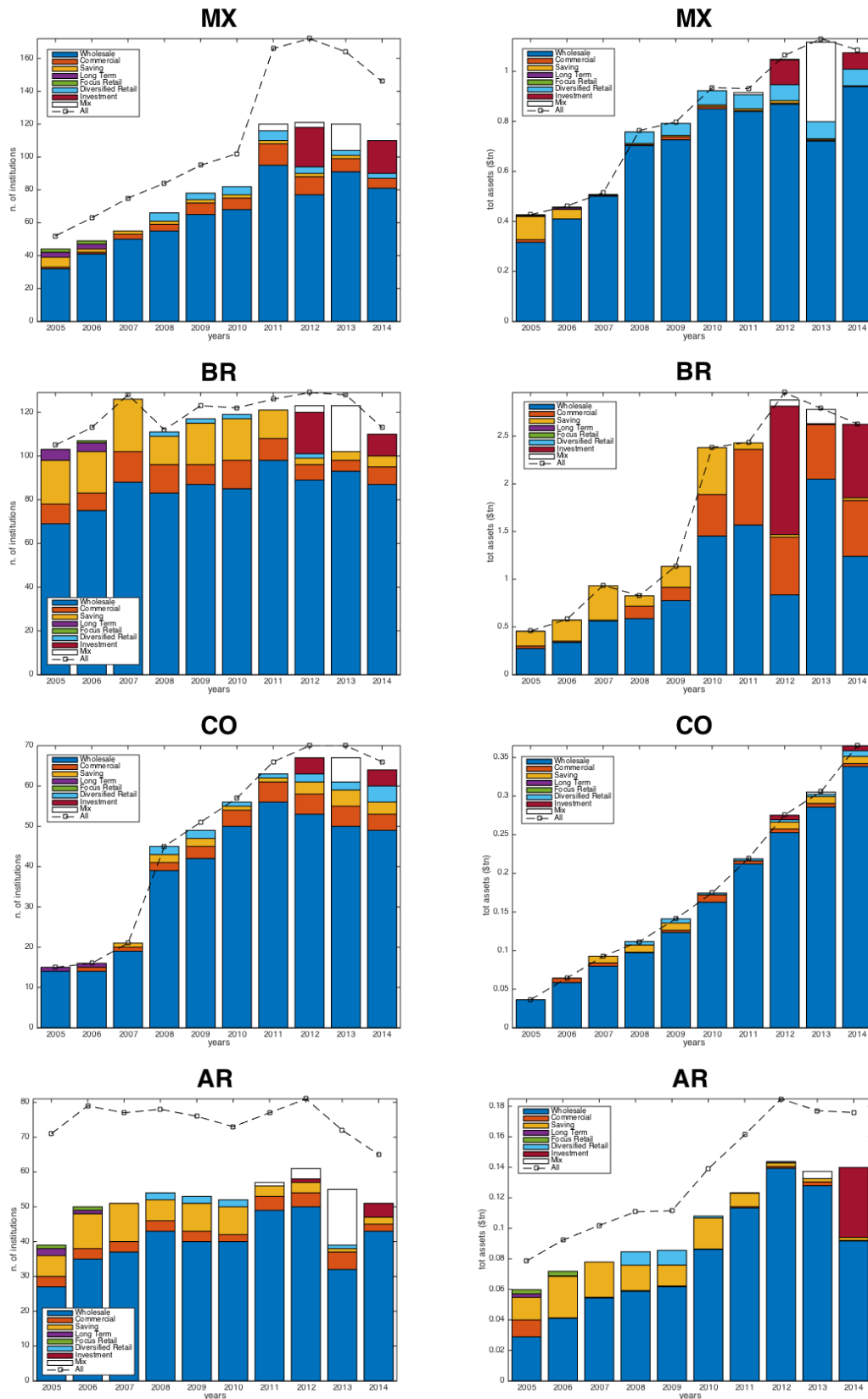
**Figure 7: Eurozone Business Models.** The plot shows the composition of banking business models in the Eurozone in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are Germany (DE), France (FR), Spain (ES), Italy (IT). Source: Bankscope, Authors' own elaborations.



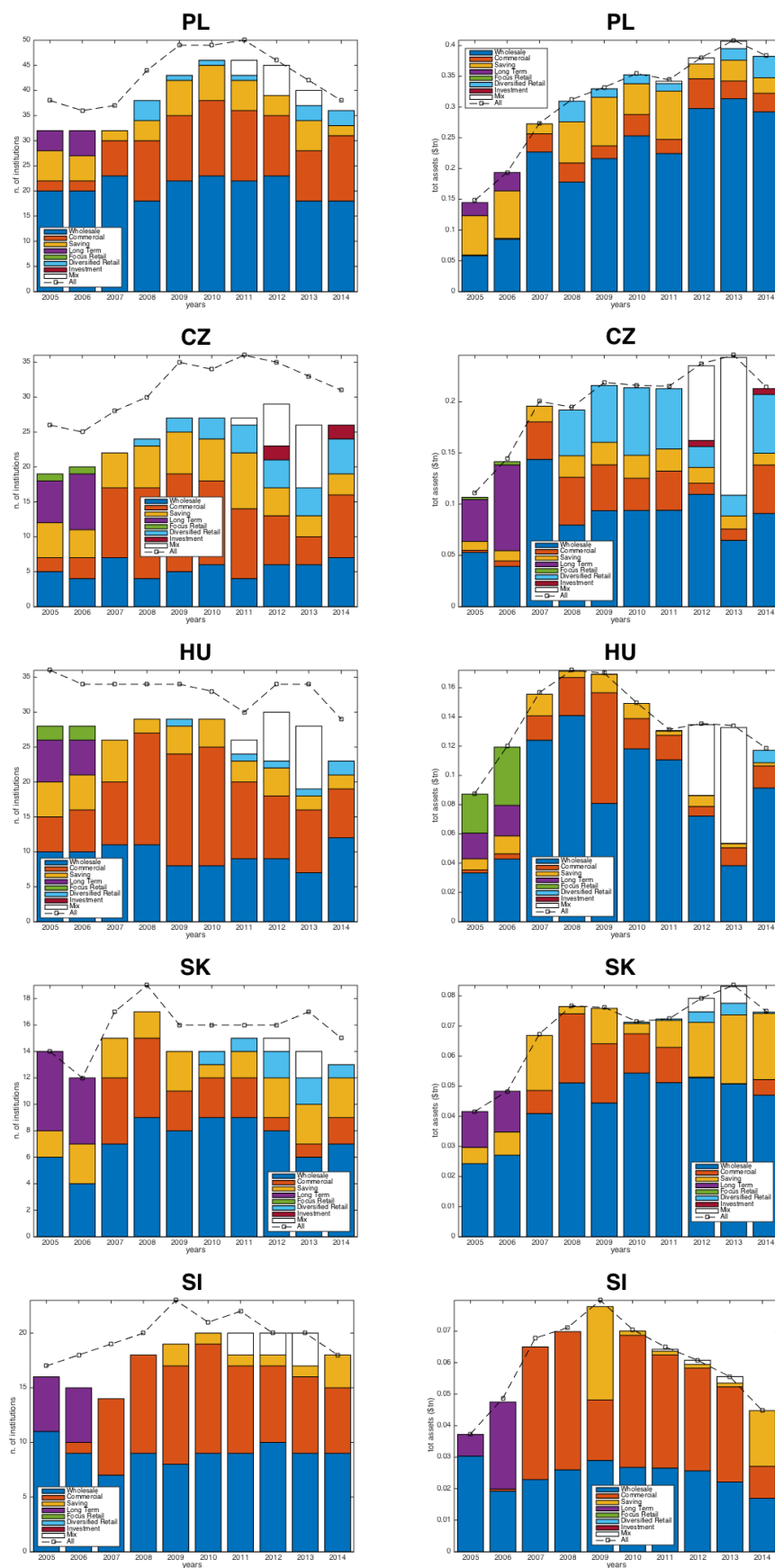
**Figure 8: Major Asian Business Models.** The plot shows the composition of banking business models in the Asian area in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are China (CN), Honk Kong (HK), Japan (JP), India (IN). Source: Bankscope, Authors' own elaborations.



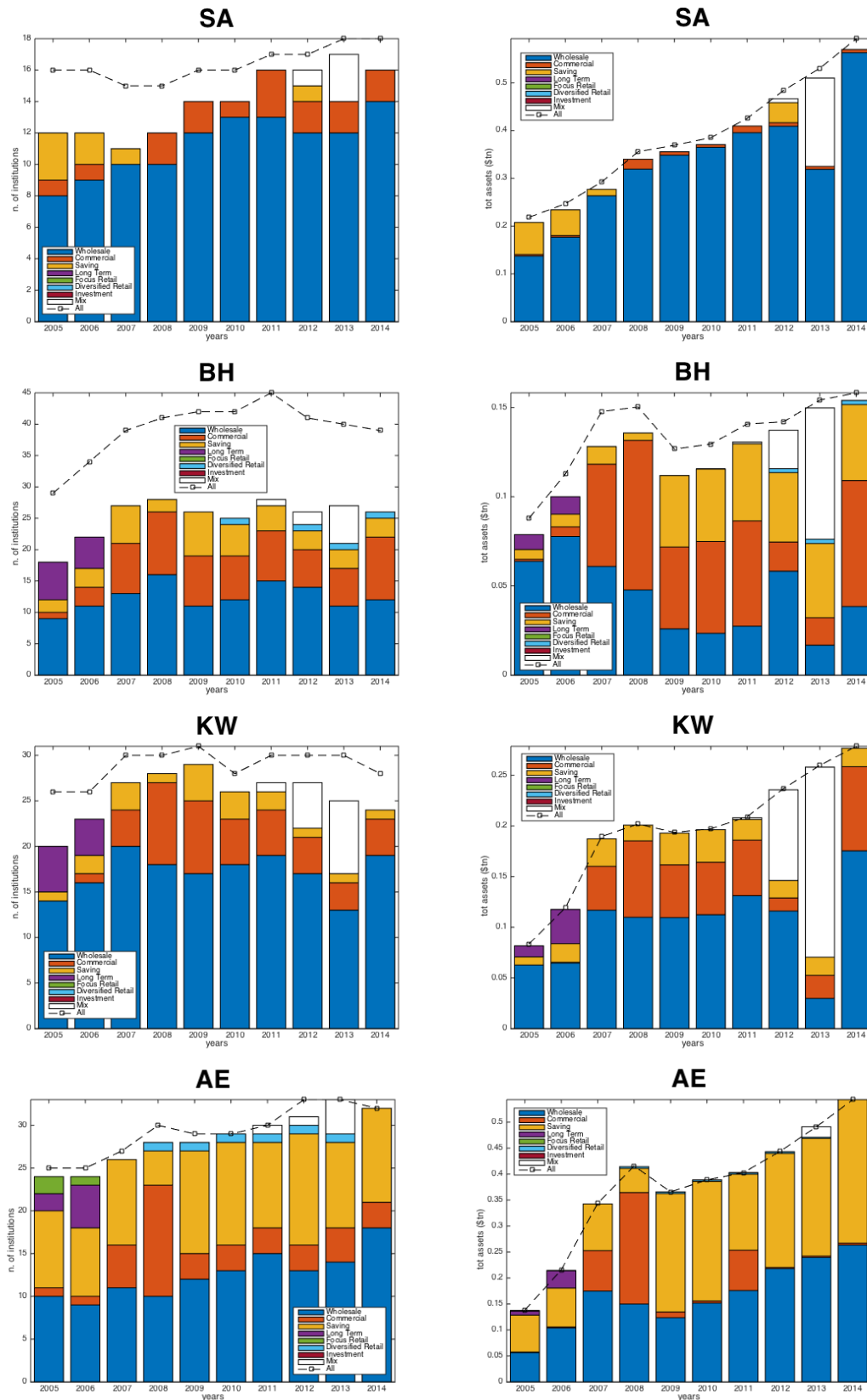
**Figure 9: Major Latin America Business Models.** The plot shows the composition of banking business models in the Latin America area in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are Mexico (MX), Brazil (BR), Colombia (CO), Argentina (AR). Source: Bankscope, Authors' own elaborations.



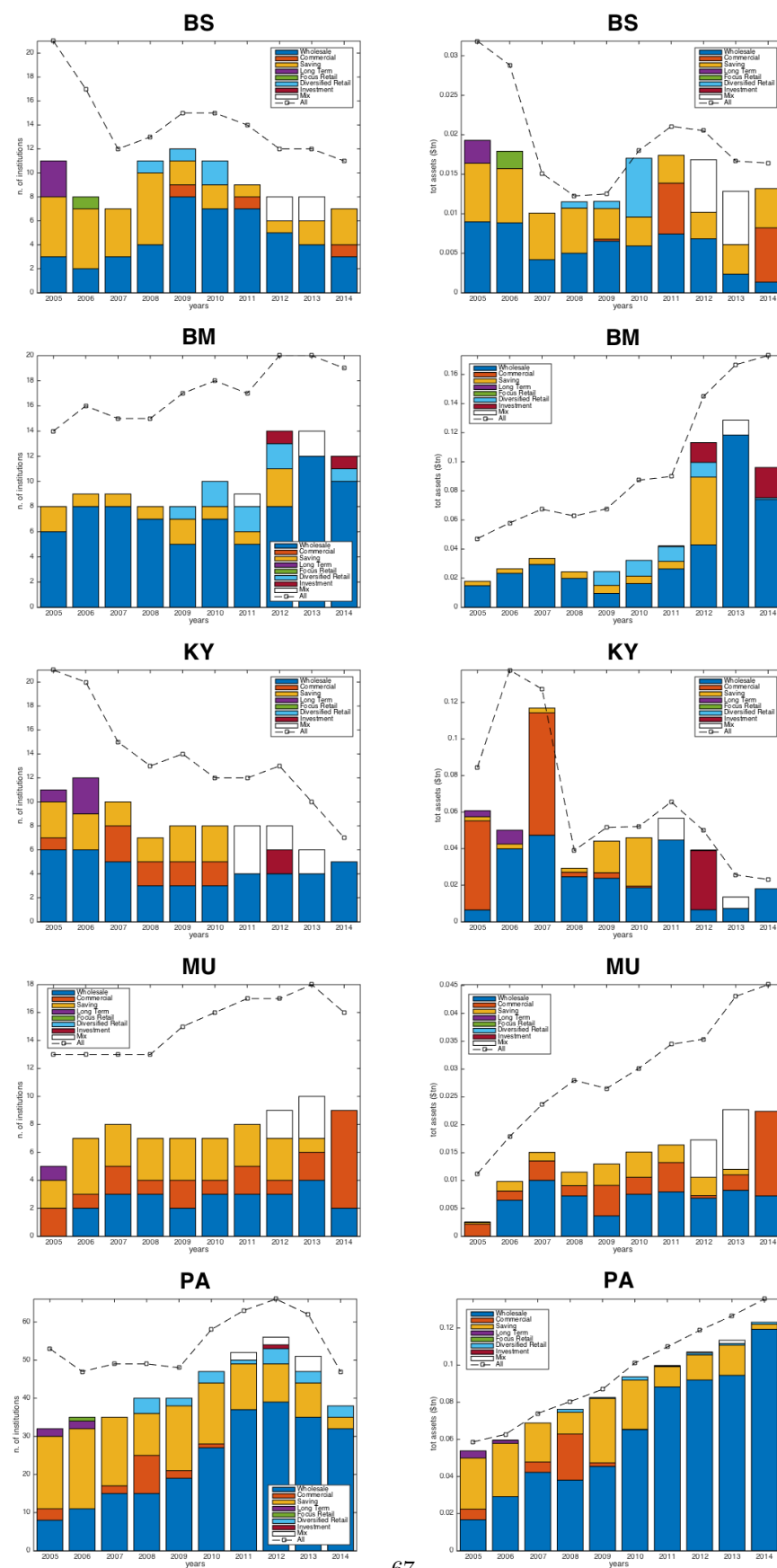
**Figure 10: Major East Europe Business Models.** The plot shows the composition of banking business models in the East Europe area in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are Poland (PL), Czech Republic (CZ), Hungary (HU), Slovakia (SK), Slovenia (SI). Source: Bankscope, Authors' own elaborations.



**Figure 11: Major Islamic Countries Business Models.** The plot shows the composition of banking business models in the Islamic area in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are Saudi Arabia (SA), Bahrain (BH), Kuwait (KW), United Arab Emirates (AE). Source: Bankscope, Authors' own elaborations.



**Figure 12: Major Offshore Financial Centres Business Models.** The plot shows the composition of banking business models within offshore financial centres in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope data set. Countries are Bahamas (BS), Bermuda (BM), Cayman Islands (KY), Mauritius (MU), Panama (PA). Source: Bankscope, Authors' own elaborations.



## E Robustness Checks

Table 25 reports the corresponding estimates of Table 11 where we simply admit the presence of one missing values in the computation of the average values used for the regressors. We rely on this basic alternative specification just to verify whether the overall framework might be influenced by the computation of regressor values. Results point to very similar estimates for each variable and model.

**Table 25: Transition Models prior to the Crisis with Proxies.** The first three columns refer to estimates computed for each peer group separately. We consider pooled observations, that means we join together the transitions between 2005 and 2006 with those from 2006 to 2007. Peer group names reported in columns' names refer to the membership before the transitions. For instance Wholesale refers to those institutions which were in that group in 2005, whose regressors are computed on average values on the interval 2003-05, and to those institutions which were Wholesale in 2006, whose corresponding regressors are averaged over the interval 2004-06. Model 1 and Model 2 include all the institutions regardless their peer group, while models with asterisks refer to the exclusion of institutions belonging to Long Term and Focus Retail groups in the biennium 2005-07. Variable Group stands for the membership in t-1 to a specific business model (the reference level is the Wholesale group). Variable G-Score refers to the year prior to the transition. We admit the presence of one missing value in the computation of average values for regressors. Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Model 1	Model 1*	Model 2	Model 2*
Intercept	-2.009*** (0.395)	-5.716*** (0.671)	-2.279*** (0.849)	1.340*** (0.147)	-3.542*** (0.269)	-2.281*** (0.202)	-3.682*** (0.276)
Total Assets	0.489 (0.400)	0.841 (0.600)	-0.919 (1.509)	-0.523** (0.243)	0.568* (0.322)	-0.175 (0.327)	0.612* (0.322)
Equity/Total Assets	-0.005 (0.005)	-0.020* (0.012)	0.019* (0.011)	-0.005** (0.003)	-0.010*** (0.004)	-0.004 (0.003)	-0.009** (0.004)
Deposits ad ST Funding/Total Funding	0.436 (0.363)	1.976*** (0.655)	-2.295*** (0.847)	-3.035*** (0.140)	0.141 (0.255)	-0.750*** (0.188)	0.206 (0.261)
Liquidity Assets/Deposits and ST Funding	0.003*** (0.001)	0.016*** (0.003)	0.010*** (0.003)	-0.001** (0.000)	0.003*** (0.001)	0.002*** (0.001)	0.003*** (0.001)
GDP per capita	-0.143*** (0.034)	0.233*** (0.048)	0.096** (0.040)	-0.012 (0.013)	-0.028 (0.021)	0.014 (0.017)	-0.007 (0.022)
Market Index	-0.022*** (0.004)	-0.006 (0.008)	0.017*** (0.004)	-0.019*** (0.002)	-0.003 (0.003)	-0.017*** (0.003)	-0.005* (0.003)
G-Score	0.090 (0.244)	1.577*** (0.180)	1.119*** (0.258)	0.741*** (0.082)	1.560*** (0.105)	1.137*** (0.096)	1.546*** (0.106)
Group(t-1): Commercial						0.296*** (0.097)	0.252** (0.099)
Group(t-1): Saving						-0.060 (0.101)	-0.044 (0.102)
Group(t-1): Long Term						4.084*** (0.094)	
Group(t-1): Focus Retail						2.706*** (0.105)	
Num. obs.	4425	3088	2946	12668	10459	12668	10459
Num. Switch	323	270	203	2413	796	2413	796
McFadden's Pseudo R <sup>2</sup>	0.040	0.145	0.081	0.061	0.047	0.372	0.047
McFadden's Adjusted Pseudo R <sup>2</sup>	0.032	0.135	0.069	0.060	0.044	0.370	0.043
LR Test	127.351***	291.872***	148.825***	805.704***	304.596***	4629.172***	314.286***

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$



To investigate the risk of distress across different business models we present in Table 26 the same models as shown in Table 12 but where we admit the presence of one missing value in the computation of average values for regressors and we replace missing values for macro and sectoral variables with geographical aggregated proxies. Results are in line with those shown in Table 12, with estimates slightly more significant for macro and sectoral regressors and less for CAMELS.

Although we are aware that a proper econometric analysis for *Long Term* and *Focus Retail* groups should consider a smaller list of regressors, we still estimate the model for these groups to provide comparisons with the *Wholesale* and *Saving* core business models. We also add the *Commercial* model whose few distress events, mainly concentrated in US, makes the estimation not well defined. For these institutions we also exploit additional data from FDIC to compute the regressors and enlarge the list of distressed institutions.

*Long Term* group has a negative and significant coefficient for net interest margin, while the coefficient for total assets is positive and significant; for *Focus Retail* institutions we observe less consistency with the other wholesale-oriented models and stronger roles for control variables. We remark that due to the small number of observations and the absence of a sufficient set of distressed institutions, results for the last two models should be taken with caution. As for Table 12, the inclusion of these marginal groups is also tested in the comparison of the last two models. Even in this approximated scenario, estimates are quite coherent between the *Wholesale-oriented\** model and the *Wholesale-oriented\* (Restricted)* model.

**Table 26: Distress Assessment within Peer Groups using Proxies.** Columns *Wholesale\** and *Saving\** refer to observations for institutions that do not change peer group and that belong to that specific business model indicated in the name of the column. Column *Wholesale-oriented\** refers in particular to institutions belonging to groups Wholesale, Commercial, Long Term and Focus Retail, admitting for transitions across these groups and never being in the Saving group in the period 2005-07 (in the case of Long Term and Focus Retail the interval that is considered is 2005-06). Column *Wholesale-oriented\* (Restricted)* is circumscribed to Wholesale and Commercial models. Column *All\** includes all institutions. For regressors definitions see Section 4.3. Asterisks stands for model specification where we admit the presence of one missing value in the computation of average values for regressors and we replace missing values for macro and sectoral variables with geographical aggregated proxies; in addition, we also exploit additional data from FDIC for distressed institutions. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimensions. Total Assets are in USD Trillion.

	All*	Wholesale*	Commercial*	Saving*	Long Term*	Focus Retail*	Wholesale-oriented*	Wholesale-oriented* (Restricted)
Intercept	-1.722*** (0.436)	-0.245 (1.053)	-0.968 (1.956)	3.957*** (1.452)	-2.560** (1.056)	-4.901* (2.790)	-2.338*** (0.476)	-1.672*** (0.614)
<sup>C</sup> Capital	0.011 (0.010)	0.006 (0.013)	0.034 (0.050)	-0.102* (0.052)	-0.003 (0.065)	-0.217 (0.295)	0.010 (0.010)	0.014 (0.010)
<sup>A</sup> Roa	-0.226*** (0.074)	-0.129 (0.091)	-0.380 (0.249)	0.642** (0.284)	-0.257 (0.318)	-0.044 (2.616)	-0.274*** (0.077)	-0.307*** (0.077)
<sup>M</sup> Roe	0.038*** (0.009)	0.046*** (0.015)	0.052 (0.040)	-0.040** (0.018)	0.000 (0.023)	0.111 (0.137)	0.046*** (0.010)	0.065*** (0.013)
<sup>E</sup> Net Interest Margin	-0.035 (0.042)	-0.027 (0.045)	0.147*** (0.056)	0.047 (0.168)	-0.791*** (0.239)	0.168 (0.427)	-0.049 (0.046)	0.013 (0.009)
<sup>L</sup> Deposits to Total Funding	-0.015*** (0.004)	-0.020*** (0.006)	0.024 (0.020)	-0.061*** (0.015)	0.004 (0.012)	-0.023* (0.014)	-0.009** (0.004)	-0.015*** (0.005)
<sup>S</sup> Total Securities to Total Assets	-0.008 (0.006)	0.002 (0.007)	0.022 (0.022)	-0.009 (0.012)	0.023 (0.014)	-0.052 (0.054)	0.001 (0.006)	-0.001 (0.006)
GDP per capita	-0.115 (0.084)	-0.039 (0.148)	-0.966*** (0.350)	-1.471*** (0.548)	-0.093 (0.210)	0.063 (0.574)	-0.065 (0.083)	-0.150 (0.117)
Inflation	0.329*** (0.090)	0.345* (0.205)	1.962*** (0.530)	0.494 (0.304)	0.845*** (0.309)	-2.820* (1.633)	0.348*** (0.103)	0.330** (0.130)
Gvt. Long-Term Yield	0.001 (0.084)	-0.477* (0.267)	-0.905*** (0.305)	0.221 (0.262)	-0.091 (0.190)	2.705** (1.174)	-0.069 (0.099)	-0.077 (0.140)
Market Index	-0.080*** (0.009)	-0.048*** (0.012)	-0.308*** (0.057)	-0.028 (0.034)	-0.097*** (0.030)	-0.074 (0.094)	-0.075*** (0.009)	-0.075*** (0.010)
Total Assets	2.013*** (0.331)	1.511*** (0.465)	3.141*** (1.197)	-0.021 (1.458)	2.432*** (0.828)	4.044 (3.100)	1.897*** (0.309)	1.407*** (0.377)
Num. obs.	5977	1965	1458	1255	949	350	5063	3956
Num. Distress Events	161	55	26	47	21	12	126	103
McFadden's Pseudo R <sup>2</sup>	0.133	0.078	0.743	0.169	0.228	0.315	0.129	0.145
McFadden's Adjusted Pseudo R <sup>2</sup>	0.115	0.024	0.645	0.100	0.091	0.050	0.107	0.117

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

# Supplementary Material

## F Non-Parametric Analysis of Peer Groups

### F.1 Multiple Comparisons

In this Appendix we show the results of the Kruskal-Wallis tests and the corresponding outcomes of the post-hoc multiple pairwise comparisons (i.e. the Dunn Test). We report for each year<sup>46</sup>: the number of institutions belonging to each group and both the values of the median and the mean of the balance sheet measures for each group. The number of asterisks (\*, \*\*, \*\*\*, \*\*\*\*, \*\*\*\*\*) in column *All* refers to the statistical difference of any given group from that number of other groups for that measure/year. *Core* stands for comparisons among institutions belonging to the 2nd and 3rd quantiles of the respective distributions. We use the Bonferroni correction to take into account the FWER. We use a restrictive significance level of 1% to judge whether two groups differ or not. Below, for simplicity we refer to groups according to the following abbreviations: I (Wholesale), II (Commercial), III (Saving), IV (Long Term), V (Focus Retail), VI (Diversified Retail), VII (Investment), VIII (Mixed; volatile group not stable in time).

Retail Loans						Retail Loans					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.00	0.15	****	****	I	3390	0.00	0.14	****	****
II	2592	0.00	0.01	****	****	II	2568	0.00	0.01	****	****
III	1946	0.00	0.02	****	****	III	1952	0.00	0.02	****	****
IV	1416	0.00	0.01	****	****	IV	1534	0.00	0.01	****	****
V	606	0.76	0.70	****	****	V	463	0.81	0.76	****	****
Retail Loans						Retail Loans					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.00	0.20	**	**	I	3270	0.00	0.11	****	****
II	3861	0.00	0.01	**	**	II	3003	0.00	0.02	****	****
III	2070	0.00	0.04	**	**	VI	2053	0.34	0.38	****	****
						III	1708	0.00	0.04	****	****
Retail Loans						Retail Loans					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.00	0.13	****	****	I	3484	0.00	0.13	****	****
II	2877	0.00	0.02	****	****	II	2760	0.00	0.02	****	****
VI	2195	0.28	0.32	****	****	VI	2233	0.29	0.32	****	****
III	2125	0.00	0.03	****	****	III	2009	0.00	0.03	****	****
Retail Loans						Retail Loans					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.01	0.16	****	****	I	3626	0.06	0.20	****	****
II	2692	0.00	0.01	****	****	VI	2244	0.30	0.34	****	****
VI	2252	0.30	0.34	****	****	II	2125	0.00	0.01	****	****
III	1778	0.00	0.03	****	****	III	1901	0.00	0.03	****	****
VIII	573	0.00	0.05	****	****	VIII	1053	0.00	0.05	****	****
						VII	310	0.00	0.02	****	****
Retail Loans						Retail Loans					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.01	0.13	****	****	I	3653	0.06	0.19	****	****
VI	2218	0.30	0.34	****	****	II	2390	0.00	0.02	****	****
II	2135	0.00	0.02	****	****	VI	2086	0.32	0.35	****	****
III	1849	0.00	0.04	****	****	III	1807	0.00	0.04	****	****
VIII	1424	0.06	0.26	****	****	VII	226	0.00	0.01	****	****

<sup>46</sup>We exclude from the non-parametric tests and from the multiple pair-wise comparisons institutions belonging to very small group or that are even singleton. The number of discarded institutions are: 28 in 2005, 11 in 2006, 0 in 2007, 10 in 2008, 3 in 2009, 1 in 2010, 0 in 2011, 2 in 2012, 0 in 2013 and 1 in 2014.

Corp. and Other Loans						Corp. and Other Loans					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.33	0.36	****	****	I	3390	0.37	0.37	****	****
II	2592	0.61	0.59	***	****	II	2568	0.62	0.61	***	***
III	1946	0.57	0.55	****	****	III	1952	0.57	0.55	****	****
IV	1416	0.65	0.61	***	****	IV	1534	0.63	0.57	***	***
V	606	0.04	0.07	****	****	V	463	0.02	0.06	****	****
Corp. and Other Loans						Corp. and Other Loans					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.29	0.34	**	**	I	3270	0.39	0.39	***	***
II	3861	0.64	0.63	**	**	II	3003	0.67	0.65	***	***
III	2070	0.54	0.52	**	**	VI	2053	0.23	0.22	***	***
						III	1708	0.54	0.51	***	***
Corp. and Other Loans						Corp. and Other Loans					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.39	0.39	***	***	I	3484	0.36	0.36	***	***
II	2877	0.64	0.58	***	***	II	2760	0.66	0.63	***	***
VI	2195	0.26	0.27	***	***	VI	2233	0.26	0.26	***	***
III	2125	0.58	0.55	***	***	III	2009	0.56	0.53	***	***
Corp. and Other Loans						Corp. and Other Loans					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.36	0.36	****	****	I	3626	0.35	0.36	*****	*****
II	2692	0.68	0.65	****	****	VI	2244	0.28	0.28	*****	***
VI	2252	0.27	0.27	****	****	II	2125	0.67	0.66	*****	*****
III	1778	0.54	0.53	****	****	III	1901	0.55	0.54	*****	*****
VIII	573	0.11	0.13	****	****	VIII	1053	0.30	0.32	*****	***
						VII	310	0.01	0.05	*****	*****
Corp. and Other Loans						Corp. and Other Loans					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.37	0.38	****	****	I	3653	0.31	0.34	****	****
VI	2218	0.29	0.29	****	****	II	2390	0.62	0.61	****	****
II	2135	0.66	0.65	****	****	VI	2086	0.28	0.28	****	****
III	1849	0.55	0.53	****	****	III	1807	0.54	0.53	****	****
VIII	1424	0.17	0.22	****	****	VII	226	0.01	0.06	****	****
Retail and Corp. Loans						Retail and Corp. Loans					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.57	0.51	****	****	I	3390	0.58	0.52	***	***
II	2592	0.62	0.60	***	****	II	2568	0.63	0.62	***	***
III	1946	0.58	0.57	****	****	III	1952	0.58	0.57	***	***
IV	1416	0.65	0.61	***	****	IV	1534	0.63	0.58	***	***
V	606	0.81	0.78	****	****	V	463	0.85	0.82	****	****
Retail and Corp. Loans						Retail and Corp. Loans					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.61	0.55	**	**	I	3270	0.58	0.50	**	**
II	3861	0.65	0.64	**	**	II	3003	0.69	0.67	***	***
III	2070	0.57	0.55	**	**	VI	2053	0.61	0.60	**	***
						III	1708	0.58	0.56	**	**
Retail and Corp. Loans						Retail and Corp. Loans					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.57	0.52	***	***	I	3484	0.54	0.49	***	***
II	2877	0.66	0.61	***	***	II	2760	0.68	0.65	***	***
VI	2195	0.59	0.59	**	**	VI	2233	0.59	0.59	***	**
III	2125	0.60	0.57	**	**	III	2009	0.59	0.56	***	**
Retail and Corp. Loans						Retail and Corp. Loans					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.58	0.52	***	***	I	3626	0.60	0.56	*****	*****
II	2692	0.69	0.67	****	****	VI	2244	0.62	0.62	*****	*****
VI	2252	0.61	0.61	****	****	II	2125	0.69	0.67	*****	*****
III	1778	0.58	0.56	***	***	III	1901	0.59	0.57	****	*****
VIII	573	0.15	0.18	****	****	VIII	1053	0.36	0.37	*****	*****
						VII	310	0.02	0.07	*****	*****
Retail and Corp. Loans						Retail and Corp. Loans					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.56	0.50	****	****	I	3653	0.59	0.53	***	***
VI	2218	0.63	0.63	****	****	II	2390	0.64	0.62	***	***
II	2135	0.67	0.67	****	****	VI	2086	0.63	0.63	***	***
III	1849	0.59	0.57	****	****	III	1807	0.59	0.57	***	***
VIII	1424	0.48	0.48	****	****	VII	226	0.01	0.07	****	****

Total Loans						Total Loans					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.76	0.70	***	****	I	3390	0.77	0.71	***	****
II	2592	0.72	0.70	***	***	II	2568	0.73	0.72	***	***
III	1946	0.73	0.71	***	***	III	1952	0.73	0.72	***	***
IV	1416	0.77	0.74	***	****	IV	1534	0.78	0.75	***	****
V	606	0.91	0.85	****	****	V	463	0.94	0.89	****	****
Total Loans						Total Loans					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.81	0.74	**	**	I	3270	0.79	0.71	**	***
II	3861	0.76	0.75	**	**	II	3003	0.79	0.77	**	***
III	2070	0.73	0.71	**	**	VI	2053	0.75	0.73	***	***
						III	1708	0.73	0.71	***	***
Total Loans						Total Loans					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.76	0.69	**	***	I	3484	0.74	0.67	**	***
II	2877	0.77	0.73	**	***	II	2760	0.77	0.74	***	***
VI	2195	0.71	0.71	**	**	VI	2233	0.71	0.71	**	***
III	2125	0.71	0.70	**	**	III	2009	0.70	0.69	***	***
Total Loans						Total Loans					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.72	0.65	***	****	I	3626	0.73	0.69	****	****
II	2692	0.77	0.74	****	****	VI	2244	0.73	0.73	***	****
VI	2252	0.73	0.73	****	****	II	2125	0.75	0.73	****	****
III	1778	0.69	0.68	**	***	III	1901	0.70	0.69	****	****
VIII	573	0.65	0.61	***	***	VIII	1053	0.65	0.61	****	****
						VII	310	0.05	0.11	****	****
Total Loans						Total Loans					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.70	0.63	**	**	I	3653	0.72	0.67	**	**
VI	2218	0.73	0.73	***	***	II	2390	0.71	0.70	*	***
II	2135	0.73	0.73	***	***	VI	2086	0.73	0.72	***	***
III	1849	0.70	0.69	**	**	III	1807	0.70	0.69	**	****
VIII	1424	0.71	0.66	**	**	VII	226	0.05	0.10	****	****
Interbank Lending						Interbank Lending					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.11	0.19	****	****	I	3390	0.12	0.19	**	****
II	2592	0.08	0.10	****	****	II	2568	0.08	0.10	***	****
III	1946	0.10	0.14	***	***	III	1952	0.12	0.15	**	***
IV	1416	0.09	0.13	***	***	IV	1534	0.10	0.17	**	***
V	606	0.06	0.07	****	****	V	463	0.07	0.08	***	****
Interbank Lending						Interbank Lending					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.11	0.20	*	*	I	3270	0.13	0.21	***	**
II	3861	0.09	0.11	**	**	II	3003	0.07	0.09	***	***
III	2070	0.12	0.16	*	*	VI	2053	0.12	0.13	**	**
						III	1708	0.11	0.15	**	***
Interbank Lending						Interbank Lending					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.12	0.19	**	***	I	3484	0.12	0.20	***	***
II	2877	0.06	0.13	**	***	II	2760	0.05	0.09	***	***
VI	2195	0.10	0.12	**	***	VI	2233	0.09	0.12	***	***
III	2125	0.08	0.13	**	***	III	2009	0.08	0.13	***	***
Interbank Lending						Interbank Lending					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.09	0.15	***	***	I	3626	0.08	0.14	**	****
II	2692	0.05	0.08	****	****	VI	2244	0.08	0.11	***	****
VI	2252	0.10	0.12	****	****	II	2125	0.04	0.07	****	****
III	1778	0.09	0.13	***	***	III	1901	0.07	0.13	****	****
VIII	573	0.45	0.44	****	****	VIII	1053	0.18	0.25	****	****
						VII	310	0.13	0.21	***	****
Interbank Lending						Interbank Lending					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.09	0.16	**	****	I	3653	0.08	0.15	**	***
VI	2218	0.08	0.10	**	****	II	2390	0.05	0.08	****	****
II	2135	0.04	0.06	****	****	VI	2086	0.07	0.09	***	***
III	1849	0.07	0.13	***	****	III	1807	0.06	0.12	****	****
VIII	1424	0.10	0.19	***	****	VII	226	0.13	0.20	***	****

Investments						Investments					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.12	0.18	***	****	I	3390	0.11	0.18	****	****
II	2592	0.22	0.24	****	****	II	2568	0.21	0.22	****	****
III	1946	0.20	0.21	****	****	III	1952	0.19	0.21	****	****
IV	1416	0.14	0.17	***	****	IV	1534	0.14	0.17	****	****
V	606	0.05	0.10	****	****	V	463	0.02	0.07	****	****
Investments						Investments					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.09	0.16	**	**	I	3270	0.08	0.14	***	***
II	3861	0.17	0.18	**	**	II	3003	0.13	0.15	***	***
III	2070	0.19	0.21	**	**	VI	2053	0.19	0.21	**	**
						III	1708	0.19	0.20	**	**
Investments						Investments					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.10	0.16	***	***	I	3484	0.12	0.17	***	***
II	2877	0.13	0.17	***	***	II	2760	0.13	0.17	***	***
VI	2195	0.23	0.24	***	***	VI	2233	0.24	0.24	***	***
III	2125	0.19	0.20	***	***	III	2009	0.20	0.21	***	***
Investments						Investments					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.12	0.17	**	****	I	3626	0.11	0.16	****	****
II	2692	0.14	0.16	**	****	VI	2244	0.23	0.23	****	****
VI	2252	0.22	0.22	***	****	II	2125	0.15	0.17	****	****
III	1778	0.20	0.21	***	****	III	1901	0.20	0.21	****	****
VIII	573	0.11	0.17	**	****	VIII	1053	0.13	0.17	****	****
						VII	310	0.35	0.38	****	****
Investments						Investments					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.13	0.19	****	****	I	3653	0.11	0.15	****	****
VI	2218	0.23	0.23	****	****	II	2390	0.18	0.21	***	****
II	2135	0.16	0.19	****	****	VI	2086	0.23	0.23	****	****
III	1849	0.20	0.21	****	****	III	1807	0.20	0.21	***	****
VIII	1424	0.11	0.15	****	****	VII	226	0.45	0.47	****	****
Customer Deposits						Customer Deposits					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.44	0.42	***	***	I	3390	0.33	0.37	****	****
II	2592	0.66	0.48	****	***	II	2568	0.65	0.48	***	****
III	1946	0.82	0.79	****	****	III	1952	0.82	0.79	****	****
IV	1416	0.57	0.57	****	****	IV	1534	0.58	0.58	***	***
V	606	0.58	0.43	***	****	V	463	0.63	0.54	**	***
Customer Deposits						Customer Deposits					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.41	0.40	**	**	I	3270	0.30	0.36	***	***
II	3861	0.60	0.49	**	**	II	3003	0.46	0.41	***	***
III	2070	0.81	0.78	**	**	VI	2053	0.72	0.66	***	***
						III	1708	0.83	0.80	***	***
Customer Deposits						Customer Deposits					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.29	0.35	**	***	I	3484	0.36	0.38	***	***
II	2877	0.41	0.37	**	***	II	2760	0.41	0.36	***	***
VI	2195	0.74	0.71	***	***	VI	2233	0.75	0.72	***	***
III	2125	0.81	0.78	***	***	III	2009	0.82	0.79	***	***
Customer Deposits						Customer Deposits					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.36	0.39	****	****	I	3626	0.45	0.41	****	****
II	2692	0.38	0.34	****	****	VI	2244	0.75	0.72	****	****
VI	2252	0.75	0.72	****	****	II	2125	0.14	0.24	****	****
III	1778	0.82	0.79	****	****	III	1901	0.83	0.80	****	****
VIII	573	0.71	0.61	****	****	VIII	1053	0.76	0.72	****	****
						VII	310	0.00	0.06	****	****
Customer Deposits						Customer Deposits					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.36	0.38	****	****	I	3653	0.53	0.46	****	****
VI	2218	0.76	0.73	****	****	II	2390	0.40	0.37	****	****
II	2135	0.23	0.27	****	****	VI	2086	0.76	0.74	****	****
III	1849	0.83	0.80	****	****	III	1807	0.83	0.80	****	****
VIII	1424	0.71	0.64	****	****	VII	226	0.00	0.06	****	****

Interbank Borrowing						Interbank Borrowing					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.16	0.24	****	****	I	3390	0.19	0.26	****	****
II	2592	0.12	0.17	****	****	II	2568	0.12	0.17	****	****
III	1946	0.02	0.06	****	****	III	1952	0.02	0.06	****	****
IV	1416	0.03	0.08	***	***	IV	1534	0.03	0.07	****	****
V	606	0.05	0.09	***	***	V	463	0.08	0.11	***	****
Interbank Borrowing						Interbank Borrowing					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.15	0.24	**	**	I	3270	0.19	0.25	***	***
II	3861	0.09	0.14	**	**	II	3003	0.04	0.14	***	**
III	2070	0.03	0.06	**	**	VI	2053	0.11	0.13	***	***
						III	1708	0.02	0.05	***	**
Interbank Borrowing						Interbank Borrowing					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.17	0.24	***	***	I	3484	0.15	0.23	***	***
II	2877	0.02	0.14	***	***	II	2760	0.04	0.15	***	**
VI	2195	0.14	0.15	***	***	VI	2233	0.13	0.14	***	***
III	2125	0.03	0.06	***	***	III	2009	0.02	0.06	***	**
Interbank Borrowing						Interbank Borrowing					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.14	0.21	****	****	I	3626	0.14	0.21	*****	*****
II	2692	0.11	0.19	***	****	VI	2244	0.12	0.13	****	*****
VI	2252	0.12	0.13	****	****	II	2125	0.17	0.22	****	*****
III	1778	0.02	0.06	****	****	III	1901	0.02	0.05	*****	*****
VIII	573	0.06	0.17	***	****	VIII	1053	0.04	0.08	*****	*****
						VII	310	0.36	0.38	*****	*****
Interbank Borrowing						Interbank Borrowing					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.15	0.22	****	****	I	3653	0.11	0.17	***	***
VI	2218	0.10	0.12	***	****	II	2390	0.11	0.18	***	****
II	2135	0.16	0.21	***	****	VI	2086	0.10	0.12	**	***
III	1849	0.02	0.05	****	****	III	1807	0.02	0.05	****	****
VIII	1424	0.05	0.11	****	****	VII	226	0.43	0.44	****	****
Long-Term Funding						Long-Term Funding					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.03	0.09	***	***	I	3390	0.04	0.10	****	****
II	2592	0.01	0.04	***	****	II	2568	0.01	0.04	****	****
III	1946	0.00	0.03	***	****	III	1952	0.00	0.03	****	****
IV	1416	0.14	0.18	****	****	IV	1534	0.11	0.17	***	****
V	606	0.06	0.10	***	***	V	463	0.11	0.14	***	****
Long-Term Funding						Long-Term Funding					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.04	0.11	**	**	I	3270	0.04	0.11	**	**
II	3861	0.03	0.11	**	**	II	3003	0.03	0.13	*	***
III	2070	0.00	0.03	**	**	VI	2053	0.03	0.07	**	**
						III	1708	0.00	0.03	***	***
Long-Term Funding						Long-Term Funding					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.05	0.12	***	***	I	3484	0.04	0.11	***	***
II	2877	0.01	0.12	**	***	II	2760	0.01	0.13	***	**
VI	2195	0.02	0.05	**	***	VI	2233	0.01	0.04	***	**
III	2125	0.00	0.03	***	***	III	2009	0.00	0.03	***	***
Long-Term Funding						Long-Term Funding					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.04	0.12	****	****	I	3626	0.04	0.12	*****	*****
II	2692	0.02	0.11	****	***	VI	2244	0.01	0.04	****	****
VI	2252	0.01	0.04	****	***	II	2125	0.02	0.12	****	****
III	1778	0.00	0.02	****	****	III	1901	0.00	0.02	*****	*****
VIII	573	0.00	0.02	****	****	VIII	1053	0.00	0.01	*****	*****
						VII	310	0.01	0.05	***	***
Long-Term Funding						Long-Term Funding					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.03	0.12	****	****	I	3653	0.04	0.12	****	****
VI	2218	0.01	0.04	***	****	II	2390	0.01	0.09	***	**
II	2135	0.02	0.11	****	****	VI	2086	0.00	0.04	***	***
III	1849	0.00	0.02	****	***	III	1807	0.00	0.02	****	****
VIII	1424	0.00	0.06	***	***	VII	226	0.02	0.05	**	***

LT Fund. and Equity						LT Fund. and Equity					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.19	0.26	****	****	I	3390	0.21	0.28	***	****
II	2592	0.08	0.10	****	***	II	2568	0.08	0.11	****	****
III	1946	0.10	0.13	***	****	III	1952	0.10	0.13	****	****
IV	1416	0.27	0.30	****	****	IV	1534	0.25	0.29	***	****
V	606	0.10	0.13	***	***	V	463	0.15	0.18	****	****
LT Fund. and Equity						LT Fund. and Equity					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.20	0.22	**	**	I	3270	0.22	0.30	***	***
II	3861	0.12	0.19	**	**	II	3003	0.15	0.23	***	***
III	2070	0.10	0.13	**	**	VI	2053	0.10	0.13	**	**
						III	1708	0.09	0.12	**	**
LT Fund. and Equity						LT Fund. and Equity					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.24	0.32	***	***	I	3484	0.22	0.30	***	***
II	2877	0.14	0.22	***	***	II	2760	0.16	0.23	***	***
VI	2195	0.09	0.11	***	***	VI	2233	0.09	0.11	***	***
III	2125	0.11	0.13	***	***	III	2009	0.11	0.13	***	***
LT Fund. and Equity						LT Fund. and Equity					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.22	0.31	****	****	I	3626	0.22	0.32	*****	*****
II	2692	0.15	0.22	****	****	VI	2244	0.10	0.12	****	****
VI	2252	0.09	0.11	****	****	II	2125	0.16	0.22	****	****
III	1778	0.10	0.12	****	****	III	1901	0.10	0.12	****	****
VIII	573	0.12	0.15	****	****	VIII	1053	0.12	0.15	****	****
						VII	310	0.13	0.17	***	****
LT Fund. and Equity						LT Fund. and Equity					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.21	0.31	****	****	I	3653	0.20	0.29	****	****
VI	2218	0.10	0.12	***	***	II	2390	0.14	0.19	***	***
II	2135	0.15	0.22	***	****	VI	2086	0.10	0.12	***	****
III	1849	0.10	0.12	***	***	III	1807	0.11	0.13	***	****
VIII	1424	0.14	0.20	***	****	VII	226	0.13	0.18	***	***
Wholesale Debt						Wholesale Debt					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.26	0.33	****	****	I	3390	0.31	0.36	****	****
II	2592	0.16	0.21	***	****	II	2568	0.16	0.21	****	****
III	1946	0.05	0.09	****	****	III	1952	0.05	0.09	****	****
IV	1416	0.25	0.26	****	****	IV	1534	0.22	0.25	***	***
V	606	0.18	0.19	***	****	V	463	0.24	0.25	***	***
Wholesale Debt						Wholesale Debt					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.28	0.34	**	**	I	3270	0.30	0.35	***	***
II	3861	0.20	0.25	**	**	II	3003	0.21	0.27	**	***
III	2070	0.06	0.09	**	**	VI	2053	0.18	0.20	**	***
						III	1708	0.04	0.08	***	***
Wholesale Debt						Wholesale Debt					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.29	0.34	***	***	I	3484	0.25	0.31	***	***
II	2877	0.18	0.26	**	***	II	2760	0.21	0.27	**	***
VI	2195	0.18	0.20	**	***	VI	2233	0.16	0.18	**	***
III	2125	0.05	0.09	***	***	III	2009	0.05	0.08	***	***
Wholesale Debt						Wholesale Debt					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.22	0.31	****	****	I	3626	0.22	0.31	****	*****
II	2692	0.26	0.30	****	****	VI	2244	0.15	0.17	****	***
VI	2252	0.15	0.17	****	****	II	2125	0.33	0.34	****	***
III	1778	0.03	0.07	****	****	III	1901	0.03	0.07	****	****
VIII	573	0.07	0.19	****	****	VIII	1053	0.04	0.09	*****	****
						VII	310	0.14	0.19	****	***
Wholesale Debt						Wholesale Debt					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.21	0.30	***	****	I	3653	0.18	0.28	***	****
VI	2218	0.14	0.16	****	***	II	2390	0.19	0.27	**	***
II	2135	0.29	0.32	***	***	VI	2086	0.13	0.15	**	***
III	1849	0.04	0.07	****	****	III	1807	0.03	0.07	****	****
VIII	1424	0.08	0.16	****	****	VII	226	0.15	0.20	*	**



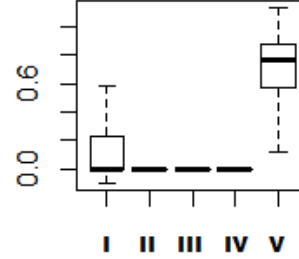
Stable Funding						Stable Funding					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.63	0.57	****	***	I	3390	0.58	0.53	***	****
II	2592	0.72	0.53	****	***	II	2568	0.71	0.53	***	****
III	1946	0.87	0.84	****	****	III	1952	0.87	0.84	****	****
IV	1416	0.83	0.78	****	****	IV	1534	0.83	0.79	****	***
V	606	0.77	0.55	****	**	V	463	0.82	0.70	****	***
Stable Funding						Stable Funding					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.64	0.61	**	**	I	3270	0.57	0.53	***	***
II	3861	0.76	0.62	**	**	II	3003	0.71	0.56	***	***
III	2070	0.87	0.84	**	**	VI	2053	0.79	0.74	***	***
						III	1708	0.88	0.85	***	***
Stable Funding						Stable Funding					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.57	0.52	***	***	I	3484	0.61	0.54	**	***
II	2877	0.66	0.52	***	***	II	2760	0.66	0.52	**	***
VI	2195	0.77	0.76	***	***	VI	2233	0.78	0.76	***	***
III	2125	0.86	0.83	***	***	III	2009	0.86	0.83	***	***
Stable Funding						Stable Funding					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.63	0.56	****	****	I	3626	0.65	0.57	*****	*****
II	2692	0.58	0.48	****	****	VI	2244	0.79	0.77	****	****
VI	2252	0.78	0.76	****	****	II	2125	0.46	0.39	****	****
III	1778	0.86	0.83	****	****	III	1901	0.86	0.84	****	****
VIII	573	0.76	0.66	****	****	VIII	1053	0.80	0.77	****	****
						VII	310	0.25	0.32	****	****
Stable Funding						Stable Funding					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.63	0.56	****	****	I	3653	0.69	0.62	****	****
VI	2218	0.79	0.78	****	***	II	2390	0.60	0.48	****	****
II	2135	0.49	0.41	****	****	VI	2086	0.79	0.78	****	****
III	1849	0.86	0.84	****	****	III	1807	0.86	0.83	****	****
VIII	1424	0.79	0.73	****	***	VII	226	0.15	0.20	****	****
Stable Funding (core)						Stable Funding (core)					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	0.55	0.50	***	***	I	3390	0.50	0.46	****	****
II	2592	0.67	0.49	***	***	II	2568	0.66	0.50	****	****
III	1946	0.84	0.81	****	****	III	1952	0.84	0.81	****	****
IV	1416	0.80	0.75	****	****	IV	1534	0.81	0.75	****	***
V	606	0.73	0.53	****	**	V	463	0.79	0.67	****	***
Stable Funding (core)						Stable Funding (core)					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	0.55	0.50	**	**	I	3270	0.48	0.46	***	***
II	3861	0.71	0.58	**	**	II	3003	0.66	0.53	***	***
III	2070	0.84	0.81	**	**	VI	2053	0.75	0.70	***	***
						III	1708	0.85	0.82	***	***
Stable Funding (core)						Stable Funding (core)					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	0.49	0.46	***	***	I	3484	0.54	0.48	***	***
II	2877	0.60	0.49	***	***	II	2760	0.60	0.48	***	***
VI	2195	0.75	0.73	***	***	VI	2233	0.76	0.74	***	***
III	2125	0.84	0.81	***	***	III	2009	0.84	0.81	***	***
Stable Funding (core)						Stable Funding (core)					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	0.56	0.50	****	****	I	3626	0.59	0.52	*****	*****
II	2692	0.53	0.44	****	****	VI	2244	0.77	0.75	****	****
VI	2252	0.76	0.74	****	****	II	2125	0.38	0.35	*****	****
III	1778	0.84	0.81	****	****	III	1901	0.84	0.82	*****	****
VIII	573	0.72	0.63	****	****	VIII	1053	0.77	0.73	****	****
						VII	310	0.05	0.11	*****	****
Stable Funding (core)						Stable Funding (core)					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	0.57	0.49	****	****	I	3653	0.63	0.57	****	****
VI	2218	0.78	0.76	****	****	II	2390	0.55	0.45	****	****
II	2135	0.42	0.37	****	****	VI	2086	0.78	0.77	****	****
III	1849	0.84	0.82	****	****	III	1807	0.84	0.81	****	****
VIII	1424	0.76	0.68	****	****	VII	226	0.04	0.11	****	****

Net Liquidity						Net Liquidity					
2005	# observations	Median	Mean	all	core	2006	# observations	Median	Mean	all	core
I	3221	-0.69	-0.62	***	***	I	3390	-0.66	-0.58	***	***
II	2592	-0.83	-0.62	****	****	II	2568	-0.82	-0.62	****	****
III	1946	-0.84	-0.81	****	****	III	1952	-0.84	-0.81	****	****
IV	1416	-0.61	-0.61	***	****	IV	1534	-0.63	-0.62	***	***
V	606	-0.70	-0.50	**	***	V	463	-0.75	-0.64	****	****
Net Liquidity						Net Liquidity					
2007	# observations	Median	Mean	all	core	2008	# observations	Median	Mean	all	core
I	4108	-0.66	-0.58	**	**	I	3270	-0.60	-0.54	**	**
II	3861	-0.75	-0.60	**	**	II	3003	-0.59	-0.51	**	**
III	2070	-0.84	-0.80	**	**	VI	2053	-0.85	-0.77	***	**
						III	1708	-0.84	-0.81	***	**
Net Liquidity						Net Liquidity					
2009	# observations	Median	Mean	all	core	2010	# observations	Median	Mean	all	core
I	3116	-0.56	-0.50	**	**	I	3484	-0.59	-0.51	**	***
II	2877	-0.54	-0.46	**	**	II	2760	-0.55	-0.46	**	***
VI	2195	-0.87	-0.84	***	***	VI	2233	-0.87	-0.83	***	***
III	2125	-0.81	-0.78	***	***	III	2009	-0.81	-0.78	***	***
Net Liquidity						Net Liquidity					
2011	# observations	Median	Mean	all	core	2012	# observations	Median	Mean	all	core
I	3608	-0.57	-0.49	****	****	I	3626	-0.59	-0.52	*****	*****
II	2692	-0.61	-0.48	****	****	VI	2244	-0.87	-0.83	*****	*****
VI	2252	-0.87	-0.82	****	****	II	2125	-0.54	-0.41	*****	*****
III	1778	-0.82	-0.79	****	****	III	1901	-0.82	-0.79	*****	*****
VIII	573	-0.70	-0.63	****	****	VIII	1053	-0.71	-0.63	*****	*****
						VII	310	-0.19	-0.21	*****	*****
Net Liquidity						Net Liquidity					
2013	# observations	Median	Mean	all	core	2014	# observations	Median	Mean	all	core
I	3424	-0.56	-0.49	***	***	I	3653	-0.59	-0.52	***	****
VI	2218	-0.87	-0.83	****	****	II	2390	-0.65	-0.49	***	****
II	2135	-0.58	-0.43	***	***	VI	2086	-0.87	-0.83	****	****
III	1849	-0.82	-0.80	****	****	III	1807	-0.81	-0.78	****	****
VIII	1424	-0.67	-0.60	****	****	VII	226	-0.26	-0.27	****	****

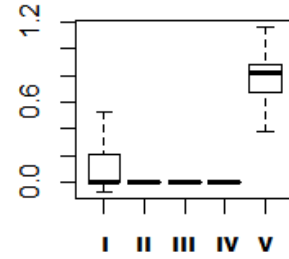
## F.2 Multiple Comparisons: examples

In this Appendix we consider some example where the distributions of balance sheet measures exhibit quite extreme patterns (in the box-plots we remove outliers for the sake of clarity in the representation). This basic investigation aims to show how the presence of groups with very peculiar distributions might influence multiple pairwise post-hoc comparisons.

Retail Loans							
2005	# obs.	Median	Mean	all	core	I-VI	I-VI (core)
I	3221	0.00	0.15	****	****	***	***
II	2592	0.00	0.01	****	***	***	**
III	1946	0.00	0.02	****	****	***	***
IV	1416	0.00	0.01	****	***	***	**
V	606	0.76	0.70	****	****		



2006	# obs.	Median	Mean	all	core	I-VI	I-VI (core)
I	3390	0.00	0.14	****	****	***	***
II	2568	0.00	0.01	****	***	***	**
III	1952	0.00	0.02	****	****	***	***
IV	1534	0.00	0.01	****	***	***	**
V	463	0.81	0.76	****	****		

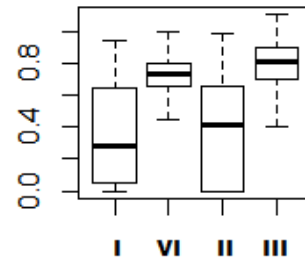
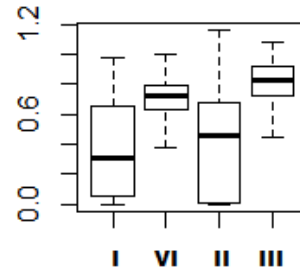


In the first example, we consider *Retail Loans* in 2005 and 2006. Distributions are quite similar in both years. Group *V* has very different median/mean values. We remove this group and we test multiple pairwise comparisons among the remaining four groups. Results confirm for the remaining groups the same statistically significant differences with or without the presence of this group.

Customer Dep.									
2008	# obs.	Median	Mean	all	core	I-II	I-II (core)	III-VI	III-VI (core)
I	3270	0.30	0.36	***	***	*	*		
II	3003	0.46	0.41	***	***	*	*		
VI	2053	0.72	0.66	***	***			*	*
III	1708	0.83	0.80	***	***			*	*

2009	# obs.	Median	Mean	all	core	I-II	I-II (core)	III-VI	III-VI (core)
I	3116	0.29	0.35	**	***				
II	2877	0.41	0.37	**	***		*		
VI	2195	0.74	0.71	***	***			*	*
III	2125	0.81	0.78	***	***			*	*



In the second example, we focus on *Customer Deposits* in 2008 and 2009. Even in this case distributions are quite similar in both years, although in 2009 groups *I* and *II* have not statistically different medians. However, we observe that group *I* and *II* are more similar between each other than the pair *III* and *VI*, and *viceversa*. Therefore, we run pairwise comparisons among this two subsets (I,II) and (III,VI). Results are in line with those obtained in Appendix F.1.

### F.3 Description of Outliers and Missing Values within Groups

We also analyse the distribution of potential outliers and missing values within each community. In particular, for each year and each community we label as outliers those institutions which fall outside a conservative range corresponding to the mean value plus or minus 1.5 standard deviations (calculated within each community separately). This analysis is computed for each variable in the vector used to determine the cosine similarities (see Section 4.1 for the list of variables). Then, we determine for each institution the outlier ratio as the proportion between measures where it has outlier values over the total number of measure in the vector (excluding the number of variables where it presents missing values). Similarly, for each institution we compute the ratio of missing values and the total ratio, where the latter is the sum of measures where it has missing values or outliers over the total number of variables used to compute the cosine similarities. Below, we report descriptive statistics for each community and year (in percentage), where values are averaged among members of each community in that specific year.

Results show that members of each community few times have outlier values in their balance sheet measures and that this pattern is quite stable over time and shared by all communities even those that emerge rarely. This is a further proof that institutions grouped in the same community have similar economic features. In addition, we consider the distribution of missing values. As emphasized in Section 5 the use of our approach is intended also to overcome issues related to the absence of some balance sheet items due to many reasons. On average the proportion of missing values does not seem to characterize any groups. This is an additional check of our approach since one might argue that the emergence of certain group can be related to the effects of missing values which clusterize specific portion of the system. We recall that the choice of the balance sheet variables included in the vector is the result of a compromise between the need to consider a wide range of measures which represent a comprehensive set of business activities and the avoidance of too specific indicators which are hardly present in many institutions' balance sheets. The intuition behind this choice is that the presence/absence of a balance sheet measure is itself a sign of a business model feature. Our results suggest that missing values are indeed well spread throughout the peer groups.

**Table 27: Outlier Ratios.** For each group and year we report summary statistics for outlier values identified as the proportion between measures where institutions have outlier values over the total number of measure in the vector (excluding the number of variables where they present missing values). We label as outliers those institutions which fall outside a conservative range corresponding to the mean value plus or minus 1.5 standard deviations (calculated within each community separately).

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>I</b>										
<i>Max.</i>	0.45	0.50	0.73	0.47	0.56	0.50	0.55	0.46	0.55	0.62
<i>Mean</i>	0.08	0.09	0.06	0.09	0.09	0.09	0.08	0.08	0.08	0.07
<i>Median</i>	0.07	0.07	0.06	0.07	0.07	0.07	0.07	0.06	0.06	0.06
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>II</b>										
<i>Max.</i>	0.67	0.64	0.50	0.50	0.60	0.45	0.55	0.42	0.45	0.55
<i>Mean</i>	0.07	0.07	0.06	0.07	0.08	0.07	0.08	0.07	0.08	0.08
<i>Median</i>	0.00	0.00	0.05	0.06	0.06	0.06	0.07	0.06	0.06	0.06
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>III</b>										
<i>Max.</i>	0.67	0.55	0.50	0.55	0.64	0.58	0.50	0.46	0.47	0.50
<i>Mean</i>	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
<i>Median</i>	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>IV</b>										
<i>Max.</i>	0.50	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.07	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>V</b>										
<i>Max.</i>	0.69	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.07	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>VI</b>										
<i>Max.</i>	0.00	0.00	0.00	0.73	0.75	0.67	0.64	0.55	0.56	0.54
<i>Mean</i>	0.00	0.00	0.00	0.08	0.08	0.08	0.09	0.08	0.09	0.09
<i>Median</i>	0.00	0.00	0.00	0.06	0.05	0.05	0.05	0.05	0.05	0.06
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>VII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.36
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.10
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.08
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>VIII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.55	0.42	0.00
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.08	0.08	0.00
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.07	0.00
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

**Table 28: Missing Values Ratios.** For each group and year we report summary statistics for missing values identified as the proportion between measures where institutions have no values over the total number of measure in the vector used to compute cosine similarities.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>I</b>										
<i>Max.</i>	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.66
<i>Mean</i>	0.45	0.45	0.44	0.44	0.42	0.42	0.40	0.38	0.38	0.40
<i>Median</i>	0.48	0.48	0.48	0.48	0.48	0.48	0.45	0.41	0.41	0.41
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
<b>II</b>										
<i>Max.</i>	0.66	0.66	0.69	0.69	0.66	0.66	0.66	0.69	0.66	0.66
<i>Mean</i>	0.50	0.50	0.49	0.49	0.50	0.49	0.48	0.47	0.47	0.48
<i>Median</i>	0.45	0.45	0.48	0.48	0.52	0.52	0.52	0.52	0.52	0.52
<i>Min.</i>	0.00	0.14	0.03	0.03	0.00	0.07	0.00	0.03	0.10	0.07
<b>III</b>										
<i>Max.</i>	0.62	0.66	0.66	0.66	0.66	0.62	0.62	0.66	0.62	0.62
<i>Mean</i>	0.43	0.42	0.42	0.40	0.38	0.37	0.36	0.34	0.34	0.35
<i>Median</i>	0.38	0.38	0.38	0.38	0.34	0.34	0.31	0.31	0.31	0.31
<i>Min.</i>	0.07	0.07	0.00	0.00	0.07	0.00	0.00	0.03	0.07	0.03
<b>IV</b>										
<i>Max.</i>	0.66	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.50	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.52	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.10	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>V</b>										
<i>Max.</i>	0.62	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.46	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.45	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>VI</b>										
<i>Max.</i>	0.00	0.00	0.00	0.69	0.62	0.66	0.62	0.66	0.66	0.62
<i>Mean</i>	0.00	0.00	0.00	0.37	0.34	0.34	0.34	0.33	0.34	0.40
<i>Median</i>	0.00	0.00	0.00	0.38	0.28	0.31	0.31	0.31	0.34	0.41
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07
<b>VII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.00	0.69
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.46
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.00	0.48
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
<b>VIII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.66	0.66	0.00
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.48	0.46	0.00
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.48	0.45	0.00
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00



**Table 29: Total (Outlier + Missing Values) Ratios.** For each group and year we report summary statistics for missing values + outlier values over the total number of measure in the vector used to compute cosine similarities.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>I</b>										
<i>Max.</i>	0.79	0.83	0.90	0.79	0.86	0.83	0.83	0.79	0.83	0.83
<i>Mean</i>	0.49	0.49	0.47	0.48	0.47	0.47	0.45	0.42	0.43	0.43
<i>Median</i>	0.48	0.52	0.48	0.52	0.48	0.52	0.48	0.45	0.45	0.45
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
<b>II</b>										
<i>Max.</i>	0.86	0.86	0.79	0.83	0.86	0.79	0.83	0.76	0.79	0.83
<i>Mean</i>	0.53	0.53	0.52	0.52	0.54	0.53	0.52	0.51	0.51	0.52
<i>Median</i>	0.48	0.48	0.48	0.52	0.55	0.55	0.55	0.55	0.55	0.55
<i>Min.</i>	0.21	0.21	0.03	0.07	0.07	0.14	0.14	0.10	0.14	0.07
<b>III</b>										
<i>Max.</i>	0.86	0.83	0.79	0.83	0.86	0.83	0.79	0.76	0.76	0.76
<i>Mean</i>	0.47	0.47	0.46	0.45	0.43	0.42	0.41	0.39	0.39	0.40
<i>Median</i>	0.45	0.45	0.45	0.41	0.38	0.38	0.34	0.34	0.34	0.34
<i>Min.</i>	0.17	0.21	0.17	0.14	0.07	0.07	0.07	0.07	0.10	0.03
<b>IV</b>										
<i>Max.</i>	0.79	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.54	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.55	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.10	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>V</b>										
<i>Max.</i>	0.83	0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Mean</i>	0.50	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Median</i>	0.48	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Min.</i>	0.17	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>VI</b>										
<i>Max.</i>	0.00	0.00	0.00	0.90	0.90	0.86	0.86	0.83	0.83	0.79
<i>Mean</i>	0.00	0.00	0.00	0.42	0.39	0.39	0.39	0.38	0.39	0.45
<i>Median</i>	0.00	0.00	0.00	0.41	0.34	0.34	0.34	0.34	0.38	0.45
<i>Min.</i>	0.00	0.00	0.00	0.14	0.17	0.14	0.17	0.17	0.14	0.14
<b>VII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.00	0.79
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.51	0.00	0.51
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.55
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
<b>VIII</b>										
<i>Max.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.83	0.76	0.00
<i>Mean</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.52	0.49	0.00
<i>Median</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.55	0.52	0.00
<i>Min.</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00



2016 © IMT School for Advanced Studies, Lucca  
Piazza San ponziano 6, 5100 Lucca, Italy  
[www.imtlucca.it](http://www.imtlucca.it)