IMT School for Advanced Studies, Lucca

Lucca, Italy

Three Essays on Systemic Risk and Financial Stability

PhD Program in Economics

XXVIII Cycle

By

Andrea Flori

2016

The dissertation of Andrea Flori is approved.

Program Coordinator: Prof. Andrea Vindigni, IMT, School for Advanced Studies Lucca

Supervisor: Prof. Fabio Pammolli, IMT, School for Advanced Studies Lucca

Supervisor: Prof. Irene Crimaldi, IMT, School for Advanced Studies Lucca

The dissertation of Andrea Flori has been reviewed by:

Prof. Sergey Buldyrev, Boston University

Prof. Giampaolo Gabbi, University of Siena and SDA Bocconi School of Management

IMT School for Advanced Studies, Lucca

2016

Life is really simple, but we insist on making it complicated

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 2 is the reproduction of the paper "*Assessing financial distress dependencies in OTC markets: a new approach using Trade Repositories data*", co-authored with M. Bonollo, I. Crimaldi, L. Gianfagna and F. Pammolli. This work partially extends a previous paper "*Systemic risk and banking regulation: some facts on the new regulatory framework*" that we published in *Corporate Ownership & Control*. The work corresponding to this chapter is forthcoming in *Financial Markets and Portfolio Management*.

Chapter 3 is the reproduction of the paper "*The Accounting Network: how financial institutions react to systemic crisis*", co-authored with A. Chessa, G. Pappalardo, M. Puliga and F. Pammolli, published in *PlosOne*.

Chapter 4 is the reproduction of the paper "*Peer-Group Detection of Banks and Resilience to Bankruptcy*", co-authored with S. Giansante and F. Pammolli.

I wish to thank my supervisor, Prof. Fabio Pammolli, for his endless and valuable source of inspiration. I wish also to thank my co-supervisor, Prof. Irene Crimaldi, for her constant support and useful advices. I received very valuable comments, that greatly improved this dissertation, from Prof. Buldyrev and Prof. Gabbi, to whom I am indebted. I also want to acknowledge my co-authors that contributed significantly to my own growth as a researcher. I am deeply in debt to a long list of post-docs and colleagues, whose support and patience greatly encouraged me during these unforgettable years at IMT.

Vita

July 11, 1986	**Born, Siena, Italy**
2008	Bachelor Degree in Economics of Financial Markets Final mark: 110/110 cum laude and *plauso accademico* University of Siena, Italy
2010	Exchange Programme (Erasmus) European Business School (EBS) Oestrich-Winkel, Germany
2010	Master of Science in Finance Final mark: 110/110 cum laude and *plauso accademico* University of Siena, Italy
2011-12	Financial Planner and Controller Cassa Depositi and Prestiti S.p.A. Rome, Italy
2015	Research Collaborator University of Siena, Department of Political Sciences Siena, Italy
2015-16	Visiting Postgraduate Scholar University of Bath - School of Management Bath, United Kingdom
2015-16	Research Collaborator Crisis Lab Project (Progetto di interesse PNR 2011-2013) Lucca, Italy

Publications

1. "Assessing financial distress dependencies in OTC markets: a new approach using Trade Repositories data," forthcoming in *Financial Markets and Portfolio Management* (with M. Bonollo, I. Crimaldi, L. Gianfagna and F. Pammolli).
2. "The Accounting Network: how financial institutions react to systemic crisis, " in *PlosOne*, October 2016 (with A. Chessa, G. Pappalardo, M. Puliga and F. Pammolli).
3. "EU ETS Facets in the Net: How Account Types Influence the Structure of the System," in Note di Lavoro Fondazione ENI Enrco Mattei, 2016-008. 2016 (with S. Borghesi)
4. "Systemic risk and banking regulation: some facts on the new regulatory framework," in *Corporate Ownership & Control*, vol. 12(2), pp. 52-63. ISSN 1727-9232. 2015 (with M. Bonollo, I. Crimaldi, M. Riccaboni and F. Pammolli)
5. "L'importanza sistemica delle banche: da una prospettiva globale a una locale? (Systemic Importance of Financial Institutions: from a Global to a Local Perspective? A Network Theory Approach)," *Bancaria*, vol. 2-2015, pp. 54-69. ISSN 0005-4623. 2015 (with M. Bonollo and I. Crimaldi)
6. "Systemic importance of financial institutions: regulations, research, open issues, proposals," *EIC working paper series*, vol. 2/2014, IMT Institute for Advanced Studies Lucca ISSN 2279-6894. 2014 (with M. Bonollo, I. Crimaldi, M. Riccaboni and F. Pammolli)

Presentations

1. "EU ETS Facets in the Net: How Account Types Influence the Structure of the System," at *Fourth IAERE Annual Conference*, Bologna, Italy, 2016.
2. "Systemic risk and banking regulation: some facts on the new regulatory framework," at *3rd Business Systems Laboratory International Symposium Advances In Business Management. Towards Systemic Approach*, Perugia, Italy, 2015.
3. "Estimating the performance and the systemic risk of a wide set of companies using financial statement data in a complex network framework," at *European Conference on Complex System*, Lucca, Italy, 2014.

Abstract

The recent financial turmoil has stressed the need to assess the risk contribution of market players to the sustainability of financial systems. Systemic risk detection is a current field of research which includes a variety of techniques borrowed from different methodologies. These approaches are designed to both evaluate how financial institutions contribute to the stability of the system and to measure how they are resilient against financial distress that might propagate throughout their connections. Cornerstone of systemic risk assessment is the representation of the system as a complex environment where agents are mutually interconnected. The aim of this thesis is to provide different perspectives on the study of systemic risk and financial stability. The first chapter introduces the assessment of co-dependencies in over-the-counter submarkets, providing a novel investigation on very detailed transactions data that have been available thanks to the new regulatory framework. Next, I exploit network theory tools to depict banking system and to analyze how the emergence of homogeneous clusters of financial institutions characterized the last decade, showing that regional communities converged to more similar banking practices after the outbreak of financial markets. Finally, I propose a rare-events logistic regression model to assess the risk of distress in a wide and globally distributed sample of financial institutions; this chapter is intended to provide evidence in support of a target risk monitoring for institutions with different business models.

Chapter 1

Introduction

The outbreak of financial systems of 2007-08 underlined the weaknesses of international regulations concerning banking risk supervision. Past regulations on banking capital requirements were focused on a *micro-prudential* approach where banks were evaluated on the basis of their financial statement structures in order to determine the corresponding regulatory capital cushions that had to be large enough to face their risk levels. Conversely, the onset of the financial crisis stressed the importance of measuring the risk contribution of each institution to the financial stability of the system, thus suggesting to replace the classical principle "too big to fail" into a "too interconnected to fail" perspective (e.g. Acemoglu et al. 2010, Acharya et al. 2009 and Hautsch et al. 2014).

Banking sector represents a cornerstone in the analysis of systemic risk due to its important role in the propagation of shocks to global markets and wider economy. As emphasized by the current crisis, banks failures weaken financial sector and spread financial distress throughout the system. Therefore, institutions whose bankrupt may trigger the default of other institutions need more rigorous supervision by regulators and should in principle fulfil higher levels of capital requirements. Hence, the need to set up an effective regulatory capital also for the systemic risk motivated the new Basel III framework to address these points (BIS 2013). Nevertheless, a huge debate over systemic risk measurement methodologies, capital requirements and effectiveness of the rules is taking place in both banking and academic communities. This leads to a growing literature concerning the identification of the key features of financial stability useful to disentangle bank's systemic risk contribution (Bisias et al. 2012, Brunnermeier and Oehmke 2012 and De Bandt et al. 2009).

One of the most promising attempts concerns the exploitation of network theory approach to build and analyse financial systems (Allen and Babus 2008, Haldane 2013, Upper 2011). Once the relationships among institutions are modeled as a network, where institutions are the nodes and their bilateral exposures are the oriented links (e.g. inter-banking transactions, OTC derivatives exposures, payment systems positions), we can exploit network theory tools and indicators to assess systemic risk and possibly identify the systemically important financial institutions (Billio et al. 2012, Boss et al. 2004, Degryse and Nguyen 2007, Elsinger et al. 2006, Haldane and May 2011, Iori et al. 2008, Langfield et al. 2014, Markose et al. 2012 and Mistrulli 2011 among others). This involves the use of simulation techniques (e.g. Eisenberg and Noe 2001 and Furfine 2003) to study cascade effects and contagion patterns to detect the configuration of the system more *resilent* to distress. In the international debate, the *resilience* of banking systems has been largely scrutinized. What happens for instance to the remaining institutions of the system when a large bank fails? Systemic features of a financial institution relate, therefore, to the losses that it can cause to the rest of the system by some *contagion* mechanisms determined by its default. This, in turns, leads to the investigation of the structures that are more prone to spread financial distress. After the pioneering works of Allen and Gale (2000) and Freixas et al. (2000), literature usually indicates two potential benchmarks: a network with a small number of large institutions with a *hub & spoke* topology or a network with a large number of medium-small institutions and a more uniform distribution of exposures. This stimulates more recent studies on the *robust-yet-fragile* configuration of the system which are pointing to both the presence of non-linearities and the need to stress complex trade-offs among risk sharing, diversification, risk monitoring and market practices (e.g. Acemoglu et al. 2010, Battiston et al. 2012, Duffie and Zhu 2011, Elliott et al. 2014, Gai and Kapadia 2010).

As emphasized by Bisias et al. (2012), literature on systemic risk detection relies on a wide range of approaches borrowed from different methodologies. Along with the complex system perspective briefly outlined above, the list of systemic risk measures includes macro-economic indicators (i.e. assets and house prices' cycles and credit-gap indicators), stress-test scores (i.e. GDP stress test or a 10-by-10-by-10 approach), cross-sectional measures (i.e. CoVarR, DIP, Co-Risk, MES) as well as forward-looking risk measures (i.e. Contingent Claim Analysis, Option iPoD, Principal Components Analysis).

Besides the introduction, this dissertation consists of three chapters, which represent individual articles. Consequently, appendices are given at the end of the corresponding chapter. Given the multidisciplinary perimeter identified by literature on systemic risk and financial stability, each of them aims to disentangle the impact of financial distress applying different perspectives and methodologies.

Chapter 2 investigates the relationships among financial sub-markets as a way to identify a potential situation of financial instability through increased co-movements among them. To study how sub-markets are mutually co-dependent, we combine the provision of granular data on *Over-the-Counter* (OTC) derivatives by trade repositories and the Joint Probability of Distress (JPoD) copula approach recently introduced by the International Monetary Fund (Segoviano and Goodhart 2009). In doing this, we define an indicator which combines some distress drivers and we observe that results on co-dependencies are close to the practical intuition: similarities between financial and contractual terms seem to be responsible for stronger co-movements among sub-markets. However, high values for JPoD even in correspondence of quite dissimilar sub-markets suggest the presence of other drivers which need to be investigated in future research.

Chapter 3 is motivated by the fact that Network Theory has been widely spotted in the latest years to study financial crisis. Since literature shows how network topology and the dynamics running on top of it can trigger the outbreak of large systemic crisis and impact on the configuration of the system, following this methodological perspective I introduce here the Accounting Network, i.e. the network can be extracted through vector similarities' techniques from banks' financial statements. Accounting Networks are built on a large database of worldwide banks in the period 2001-2013, covering the onset of the global financial crisis of mid-2007. These networks are analysed both in terms of their topological properties and for the emergence of homogeneous communities of banks. Remarkably, enough sensible regional aggregations show up with the Japanese and the US clusters dominating the community structure, although the presence of a geographically mixed community points to a gradual convergence of financial institutions to similar supranational banking practices. In the last part of the chapter, Principal Components Analysis is applied to reveal the main economic dimensions that influence communities' heterogeneity. Even using the most basic vector similarity hypotheses on the composition of the financial statements, the signature of the financial crisis clearly arises across the years around 2008.

Chapter 4 is partially a follow up of Chapter 3 since it investigates similarities of banks' business models by employing a structural clustering approach based on banks' balance sheet information. This chapter applies the same clustering methodology discussed in Chapter 3, although here we consider a wider set of institutions with slightly different available financial statement items. The resulting clusters identify major important banking peer groups characterised by the assets and liabilities structure. We performed the peer-group assessment on more than 10k banks covering more than 180 countries during the period 2005-14. Main large peer groups are quite stable over time, with a couple of main structural breaks coinciding with the introduction of new accounting standards and the 2007-08 global financial crisis. Both individual banks' membership to a particular business model and the geographic representation is investigated. The peer-group membership is then tested as explanatory variable to study banks distress events after the breakdown of 2007. The analysis confirms the importance of CAMELS (i.e. financial statements indicators) 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 model membership over time. For the first dimension, CAMELS variables along with macro and sectoral features contribute differently, sometimes with opposite sign, to the likelihood of distress among institutions with a different business model. For those institutions which 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. A bank supervisor would definitely benefit from monitoring these true business model features for a more accurate and targeted intervention in stabilizing the banking sector.

Chapter 2

Assessing Financial Distress Dependencies in OTC Markets: a New Approach using Trade Repositories Data

2.1 Introduction

The 2007-08 financial crisis was mostly due to liquidity and credit (counterparty) risks within the banking system. Although liquidity and OTC (*Over-the-Counter*) derivatives were among the main causes of distress, one of the most surprising effects was the contagion to other financial players and to other markets and sectors. Consequently, this motivated the introduction of *Systemic risk* as a new *building block* in the regulatory framework, such as the Basel III regulation (see e.g. BIS (2013) and FSB (2015)). Financial instability and systemic risk assessment are receiving a growing interest among researchers and regulators, and many different approaches and techniques have been proposed so far[1]. In particular, recent literature on financial systems[2] focused on the payments system, the interbanking deposit markets and the OTC derivatives markets. However, the latter is one of the most difficult to investigate due to the complexity of the "underlying" transactions, i.e. the derivatives' payoffs with their highly customized structures, and the scarce availability of detailed data especially for past transactions. Aggregated statistics on OTC derivatives markets are usually released by international organizations such as BIS (Bank for International Settlement) and OCC (US Office of Comptroller of the Currency), or banking associations like ISDA (International Securities and Derivatives Association). The collapse of 2007-08 stressed the need for a better provision of data in order to assess systemic risk and prevent market abuse. Therefore, changes in regulatory frameworks pointed to a more detailed description of the deals, thus depicting a more representative and updated picture of derivatives markets (see e.g. Duffie et al. (2010) and Russo (2010)). In US, prior to the Dodd-Frank Act (US 111th Congress (2010)), financial institutions had less obligations regarding the amount of financial leverage, counterparty risk exposures, market share, and other data to be reported to any regulatory agency. Conversely, new rules introduced also requirements on OTC exposures and assigned to specific agencies the role of collecting and sharing data. Similarly, in Europe the creation of the European Securities and Markets Authority (ESMA) and the European Systemic Risk Board (ESRB) were motivated also by the need to enforce the availability of data to improve the supervision and the restraint of systemic risk (EC 2013a; EC 2013b). In addition, the European Parliament established the European Market Infrastructure Regulation (EMIR) with Regulation No. 648/2012 (EUP 2012). Both the EMIR in Europe and the Dodd-Frank Act in US aim to disclose a more detailed description of the derivatives mar-

[1]We omit the review of this strand of literature, recommending the interested reader to refer to Bisias et al. (2012) and Brunnermeier and Oehmke (2012) and the references therein for a detailed analysis of financial stability measures and models used for assessing systemic risk.

[2]For instance a useful review on the application of network theory tools and methodologies can be found in Upper (2011).

kets. Although only authorities are allowed to exploit the highest level of granularity, market players can benefit from this flow of data through trade repository services (TRs), which collect and match data and allow the public access to this information[3]. In Europe this corresponds to an intermediate level where data are aggregated according to e.g. different asset classes and maturity features, while in US transaction data are reported almost in real-time and only confidential data are not available[4]. Therefore, nowadays an increasingly need for transparency is required: for what it may concern OTC derivatives and central counterparties, useful analyses are in Cecchetti et al. (2009) and Hull (2014), while for the interest rate derivatives markets some insights can be found in Avellaneda and Cont (2010) and Fleming et al. (2012).

In this work we attempt to describe how segments of the OTC derivatives market are related to each other. In particular, we focus on reciprocal co-movements during distressed market conditions using a novel database on OTC transactions which is based on trade repositories data. Hence, to study how sub-markets are mutually influenced we deal with the following issues. First, we identify a suitable set of OTC sub-markets within the IRS instruments by aggregating deals according to financial and contractual terms. We circumscribe our analysis on the most common type of IRS contracts, i.e. the *Fix-to-Floating* instruments, taking into account deals where the underlying rate is *USD-LIBOR-BBA*, contractual start is *Spot* and currency is *US Dollar*. This represents the most significant subset in our dataset (which is supplied by IASON ltd[5]). The identification of sub-markets is then driven by the *maturity* of the contract, the *frequencies* of the swap legs and the presence of *clearing* agreements. Second, we construct an indicator for assessing the level of distress present in these sub-markets. This distress indicator combines sev-

[3]For a detailed description of trade repositories activities, see e.g. DTCC (2013) and DTCC (2014).

[4]For a deeper study on the divergences between EU and US in financial markets' regulations, we suggest for instance Acharya et al. (2010), Lannoo (2013) and Valiante (2010), while a valuable reference for a better understanding of the key requirements involved in the aggregation of TRs data are provided by FSB (2014).

[5]Iason ltd is a consulting firm operating in risk management tools and applications. For references see *http://www.iasonltd.com/.*

eral dimensions useful to measure market conditions, such as a proxies for the bid-ask spread of the prices, their volatility, the number of deals and the average traded volumes. Basically, although we are aware of the fact that market distress might be related to a wide set of factors, which might present interacting effects, we prefer to focus on a simple and intuitive framework that can synthesise the main forces affecting market dynamics. Therefore, the aim of this indicator is to reflect some of the most evident and relevant dimensions that influence the ordinary course of business within OTC sub-markets. Third, we analyse the distress dependence between pairs of sub-markets by means of the copula theory and we investigate the joint distribution of the increments of their distress indicator. Copula functions provide mathematical tools for the modelization of multivariate stochastic dependence structures, that are able to capture various forms of stochastic dependence, not only linear dependencies. In particular, we estimate the Kendall's tau correlation coefficients and the joint upper-tail probabilities (henceforth Joint Probabilities of Distress). Our approach is similar to the one introduced in the IMF Banking Stability Measure report by Segoviano and Goodhart (2009) to describe the distress interdependent structure between financial institutions. However, it is worth noting that our context of application is completely different and, therefore, we need to face technical issues that are specific of our case-study (for instance, we do not have "default" thresholds and so the CIMDO methodology used in Segoviano and Goodhart (2009) is not feasible in our case).

Although our approach exploits standard methods used in risk management, to the best of our knowledge the present work is one of the first empirical studies based on micro-founded trade repositories' data. Related literature comprises e.g. Slive et al. (2012) who analyse central clearing effects in credit default swaps (CDS) markets through Intercontinental Exchange (ICE) Trust and Clear Europe data and Markose et al. 2012 who investigate the role of Systemically Important Financial Institutions (SIFIs) within US CDS market using Federal Deposit Insurance Corporation (FDIC) data. A very recent paper which exploits data from the Depository Trust and Clearing Corporation (DTCC) is Gehde-Trapp et al. (2015); however it focuses on CDS rather than on IRS. A comparison between official BIS statistics and detailed trade repositories data is in Bonollo et al. (2015) who describe how OTC derivatives market segmentation can be implemented through the provision of more granular flows of information related to the new regulatory framework. The novelties of our analysis are both the originality of the dataset that we exploit to identify specific sub-markets and the distress indicator that we introduce. We note that, despite the several difficulties to be faced due to the pioneering nature of our work (for instance the quality of the TRs data, the sub-markets identification and the new distress indicator definition), our outcomes are consistent with the practical intuition. While analysing in a micro-prudential approach the portfolio and the risks of a bank is a complex but rather sharp task, to infer from global market data the risks of a financial system as a whole portfolio is a current frontier of the research. In the past, lack of detailed data and the difficulty to converge towards an accepted definition of systemic risk (see e.g. IMF-BIS-FSB (2009) and FSB (2010)) made very challenging the measurement of distress signals arising from financial markets. This work aims to introduce in the debate on systemic risk assessment and financial stability a way to exploit trade repositories data to detect distress and crisis phenomena.

The paper is organized as follows: after a detailed description of both the dataset and the procedure applied for sub-markets identification (Section 2.2), we introduce the indicator used to investigate distress dependencies among sub-markets (Section 2.3). Then, Section 2.4 explains in detail the methodology that we have utilised to estimate codependencies. Finally, the results of our analysis are illustrated and discussed in Section 2.5. Section 2.6 concludes and makes some remarks about future lines of research.

2.2 Description of the Dataset

International statistics on OTC markets are usually provided by some organizations such as BIS and OCC, or banking associations like ISDA. These statistics are based on some main reporting dealers, for instance the largest (a few dozens) commercial and investment banks that regularly send some low granular data on their own derivatives deals to these central organizations which publish them once having applied data cleaning procedures to avoid e.g. double counting issues. Although this flow of data covers a high percentage of the global OTC markets, the information related to both asset classes and payoffs is not very detailed and may be not comparable among different data providers. Mark-to-market consensus prices may differ from pre-trade indicative prices and from the actual trade prices at which derivatives are exchanged. For these reasons, we rely on a trade repository dataset retrieved from *GTR-Analytics*[6], which collects trades' information from several trade repositories and for many types of instruments, controlling for manifest inconsistencies and mismatches. The latter process limits potential biases due to data misreporting and fragmentation which arise from merging datasets from different sources and across many regulations.

Our study focuses on the interest rates derivatives market which at the end of December 2014 accounts for the 80% and the 75% of the global OTC derivatives market in terms of the outstanding notional amount and the gross market value, respectively. Since the swaps market was $\$381$ trillion compared with $\$505$ of the total outstanding notional amount of the interest rate market[7], this motivates our choice to study the swaps segments as a representative case-study for the global OTC derivatives market. In particular, for each deal (identified by an ID) our database specifies the asset class of the instrument and reports a set of information regarding contractual terms, including for instance: the execution time, the effective date and the contractual expiry of the deal, the settlement and both the underlying assets currencies, payment frequencies, day count convention, and, obviously, the notional and the price. In addition, we can also exploit information on clearing agreements and collateral positions which enrich the description of market trends and improve risk assessment. As a matter of fact, it is worth outlining that we

[6]This is a software developed by the consulting firm IASON ltd. For references see *http://www.financial-machineries.com/gtr-analytics.htm*

[7]Data refer to BIS statistics and to single currency contracts only. For further references, see *http://www.bis.org/statistics/derstats.htm*.

refer to prices and volumes of actual traded deals in the market, which consequently extend the traditional use of *offered* rates (bid/ask quotes shown by brokers or data providers) and *consensus* data (quotes/prices submitted by market contributors).

2.2.1 Sub-Markets Identification

The identification of a robust sub-market concept is not straightforward. Along with some financial drivers that can support us for clustering the whole market, we take into account also the availability and the quality of a wide set of deals for different financial instruments. However, although the methodology we propose is still somehow heuristic, we believe that at this first stage of the study this is a reasonable approach for the analysis of OTC sub-markets' co-movements.

To ensure comparability, we circumscribe our study to *Fix-to-Floating* instruments. For the same reason, we consider contracts where the underlying rate is *USD-LIBOR-BBA*, contractual start is *Spot* and currency is *US Dollar*. This represents the most significant subset in our dataset. In particular, the identification of sub-markets is driven by the *maturity* of the contract, the *frequencies* of the swap legs and the presence of *clearing* agreements. Data investigation suggests to consider Fix-to-Floating instruments with leg frequencies equal to *(3m vs 3m)* and *(6m vs 3m)*. In addition, we aggregate deals according to three main maturities, i.e. less or equal to 2 years (*Short*), between 2 years and 10 years (*Medium*) and greater or equal to 10 years (*Long*). Finally, we distinguish contracts between those for which there are clearing agreements (*C*) from those for which uncleared (*UC*) conditions are present (see Table 1).

The frequency of the leg payments became a relevant factor after the financial crisis, when it was clear that the timelines of cash flows changed both the liquidity (funding) risk and the counterparty risk for the two involved financial agents. This is well known as the *multiple curve new framework*[8]. In other words, one cannot evaluate financial instruments without considering the frequency of cash flows since *ceteris paribus* the

[8] See Pallavicini and Brigo (2013).

Table 1: Sub-Markets Definition. Fix-to-Floating refers to contracts with swaps legs frequencies equal to *(3m vs 3m)* or *(6m vs 3m)*. *Short*, *Medium* and *Long* refer to deals with maturities less or equal to 2 years (*Short*), between 2 years and 10 years (*Medium*) and greater or equal to 10 years (*Long*). Finally, data are further partitioned according to the presence (*C*) or absence (*UC*) of clearing agreements.

Sub-mkt	**Fix-to-Floating**	**Maturity**	**Clearing**
1	*(3m vs 3m)*	*Short*	*C*
2	*(3m vs 3m)*	*Medium*	*C*
3	*(3m vs 3m)*	*Long*	*C*
4	*(6m vs 3m)*	*Short*	*C*
5	*(6m vs 3m)*	*Medium*	*C*
6	*(6m vs 3m)*	*Long*	*C*
7	*(6m vs 3m)*	*Short*	*UC*
8	*(6m vs 3m)*	*Medium*	*UC*
9	*(6m vs 3m)*	*Long*	*UC*

IRS fair values will be slightly different. Also the netting flag is a very informative variable. For instance, both the Dodd-Frank act and the ESMA regulations ask financial institutions to compulsorily apply netting agreements in the transaction management to keep as low as possible the credit exposures. In addition, even enterprises are obliged to this practice for deals above some relevant thresholds (e.g. 3 bn of Euro in terms of outstanding notional for interest rate derivatives in the ESMA regulation). For this reason, we can assume that the Yes/Not clearing agreement digit can be used as a proxy for the counterparty class, i.e. financial institutions vs. enterprises.

Although information on traded deals is available also for the first part of 2013, for the following analysis we consider only data from *September 2013* to *April 2015* since the amount and quality of reported deals at the beginning of 2013 is not satisfactory. This choice ensures a good availability of data along the reference period. Table 2 shows descriptive statistics for each sub-market[9].

[9]We apply a further check for double counting in the deals by controlling for contractual terms. We consider as duplicated deals those transactions that are equal in terms of: dissemination id, contractual expiry, effective date, end date, price and notional amount.

Table 2: Number of Deals and Volumes for each Sub-Market. Descriptive statistics refer to the number of deals and their notional amounts (in million of Us Dollar) during the whole interval from September 2013 to April 2015. In the upper part we show data for contracts with swaps legs frequencies equal to (6m vs 3m), while in the lower part we consider deals with swaps legs frequencies equal to (3m vs 3m). Short, Medium and Long refer to deals with maturities less or equal to 2 years (Short), between 2 years and 10 years (Medium) and greater or equal to 10 years (Long), respectively. Finally, contracts are further partitioned according to the presence of clearing agreements (Cleared vs. Uncleared).

IRS (3m x 3m)

	Short		*Medium*		*Long*		***Total***	
	Cleared	*Uncleared*	*Cleared*	*Uncleared*	*Cleared*	*Uncleared*	***Cleared***	***Uncleared***
Number of deals	1,170	154	9,028	706	9,406	133	**19,604**	**993**
Notional amount	175,226	8,359	874,483	34,479	412,779	4,131	**1,462,488**	**46,969**

IRS (6m x 3m)

	Short		*Medium*		*Long*		***Total***	
	Cleared	*Uncleared*	*Cleared*	*Uncleared*	*Cleared*	*Uncleared*	***Cleared***	***Uncleared***
Number of deals	14,063	2,128	100,555	9,547	92,008	9,405	**206,626**	**21,080**
Notional amount	2,301,060	294,994	9,258,744	745,467	4,134,122	396,999	**15,693,926**	**1,437,460**

Results suggest that deals involving *Fix-to-Floating* instruments with leg frequencies equal to (6m vs 3m) are more frequent than those with leg frequencies equal to (3m vs 3m). This is more evident once we consider deals characterised by the absence of clearing agreements. In particular, short maturities are less diffused, while figures are comparable for the *Medium* and the *Long* sub-sets (given a certain swap legs frequency). In order to identify sub-markets, these descriptive statistics suggest to discard un-cleared deals of *Fix-to-Floating* instruments with leg frequencies equal to (3m vs 3m) due to data limitations. Therefore, our final list of sub-markets is composed by six sub-sets with leg frequencies equal to (6m vs 3m) and three sub-sets with leg frequencies equal to (3m vs 3m), the latter characterised by the presence of clearing agreements.

2.2.2 Comparisons with other Data Sources

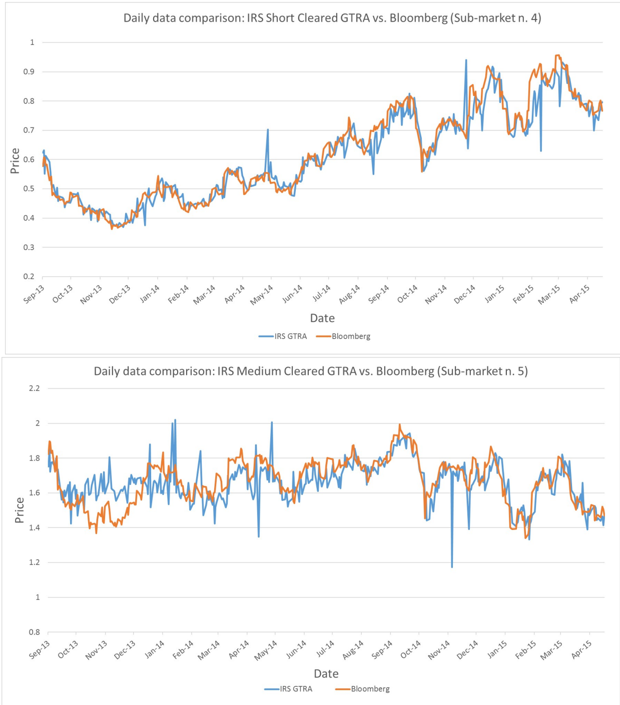
Official BIS descriptive statistics provide information for OTC derivatives by currency. At end-December 2014[10], US Dollar interest rate swaps were 124 trillion in terms of outstanding notional amount, while in our dataset (*Fix-to-Floating* 3m3m plus 6m3m) the amount is about 14.6 trillion of US Dollar, i.e. close to the 12% of the whole USD IRS market. Although a direct comparison between BIS statistics and our sample would require a more detailed partition of the deals, not yet available in the BIS statistics, however we observe a satisfactory coverage of IRS markets provided by our dataset.

In addition, in Figure 1 (top in the panel) we compare IRS prices of the short maturity bucket in the cleared case (sub-market n. 4) from the GTRA database vs. data we obtain from *Bloomberg* corresponding to the USD 2Y curve. Time series trends are very similar during the entire reference period with only few exceptions, most of which due to a sharper reported price from our data provider. Similarly, the comparison between the USD 5Y curve and the medium maturity bucket (sub-market n. 5) that is shown at the bottom of the panel of Figure 1 confirms an overall coherence among both sources. In this case we can observe some differences especially in the first period, although on average both sources of data depict a similar picture. Those differences might be due to a grouping effect, since even if the medium bucket is mainly influenced by the 5Y tenor, the presence of other maturities in the bucket (e.g. 3Y, 4Y, 7Y) may affect the aggregated level.

Finally, in Figure 2 we focus on the time series for the un-cleared case with short maturity (sub-market n. 7). We can observe a quite erratic dynamics. We recall that the "un-cleared" flag represents a signal for the fact that the counterparty is more likely to be a corporate instead of another financial institution. Therefore, different factors could determine an apparently strange behaviour of price series:

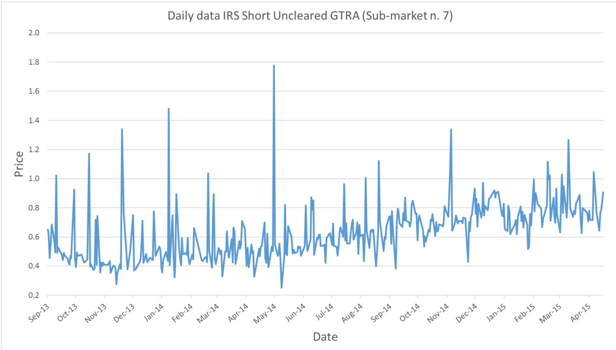
[10] For references, see statistics in *http://www.bis.org/statistics/dt07.pdf*.

Figure 1: Comparison between GTRA and Bloomberg Time Series. IRS GTRA refers to the aggregated time series for deals belonging to sub-market 4 (top panel) and sub-market 5 (bottom panel). Bloomberg curves refer to USD 2Y and USD 5Y, respectively. Prices are in percentage.



- an *up-front* is stated in the deal, i.e. one of the two counterparties receives immediately a cash amount. To balance it, the IRS fixed leg might be shifted to offset the upfront;
- the IRS pay-off could be very customized, hence it requires a different fixed level;
- there is less liquidity in the IRS segment for the enterprises. Banks may apply some relevant *mark-up* to off-set the counterparty risk;
- a combination of the above factors.

Figure 2: GTRA Price Time Series for Sub-Market n. 7. Time series refers to deals belonging to sub-market 7. Daily prices are computed according to the weighted average of the prices of the contracts, where weights are based on the notional traded amount of the deals. Prices are in percentage.



2.3 Distress indicator

Given a certain set of sub-markets representative for the global OTC market of swap instruments, we propose a way to measure their market conditions and to identify whether pairs of sub-markets are reciprocally co-dependent. In particular, we are interested in sub-markets' co-movements that point to distressed scenarios. Therefore, we are willing to introduce an indicator of distress which synthesizes several dimensions of market conditions. Before exposing how this measure is defined, it is worth stressing that we do not rely on traditional concepts of default since markets cannot go bankrupt in a strict sense, although the absence of transactions can be interpreted in a similar way.

In order to assess the level of financial distress within each sub-market, we propose an indicator of distress able to capture several aspects related to the financial stability. We assume that the main forces affecting the level of distress of a sub-market are represented by *i)* the bid-ask spread of the prices, *ii)* the volatility of the prices, *iii)* the number of deals and *iv)* the volumes of the notional traded amount. These hypotheses reflect the perception that a wider bid-ask spread stands for deteriorated liquidity conditions as well as a higher price volatility may suggest the presence of a distressed scenario. Similarly, a lower number of deals (or modest average notional traded amounts) may be a signal of slowness in the process of adjusting prices, which may impact on the capacity to close positions. In addition, these forces may present interacting effects. However, in our work we prefer to rely on a simple approach, thus focusing only on the direct contributions of the four aforementioned measures. For point *i)*, since we cannot directly deal with bid-ask quotes and we are not aware of the parts involved in the transactions, we rely on the ratio between the maximum and the minimum of daily prices as an acceptable proxy for the bid-ask spread within a certain sub-market. This choice is motivated by the fact that a tight daily deviation between the maximum and the minimum is likely to imply that traded deals have been priced within a close interval. Although our choice represents a basic approximation of the bid-ask spread, it is worth underlining that recent works on the estimation of the bid-ask spread point also to high/low prices as a way to measure bid and ask quotes in financial markets (Corwin and Schultz 2012; Deuskar et al. 2011). For point *ii)* we compute the dispersion in terms of the standard deviation of daily prices, while for points *iii)* and *iv)* we determine the daily number of deals (cardinality) and the daily average of the traded notional amounts, respectively. Finally, in order to get less noised estimates we aggregate these measures on a weekly interval[11].

To gauge the presence of distressed conditions, we recall that even in the Basel Vasiceck-Gordy model, although single default probabilities are present, the use of the 99.9% quantile for the capital charge is not related to a specific event, since it is just a very regulatory confidence level for the estimation of the losses in the credit portfolio. Therefore, it seems reasonable to avoid the selection of a given threshold above which we state that a certain sub-market experiences distress. Indeed, we suggest to analyse sub-markets' reciprocal behaviours in the tail corresponding

[11] If missing values are present due to lack of data, we replace them by the cubic spline interpolation of the available points. In order to limit potential biases due to outliers, for each sub-market we cut off 0.025 of the area in each tail of the reference sample distribution.

to detrimental conditions. Obviously, a threshold level could be set up in some further research to design for instance a proper *backtesting* procedure for the model. Finally, we recall that the IRS price level represents an average of the *forward* (expected) interest rates over the IRS maturity. Hence, any turmoil in the IRS price and/or observed volumes could jointly reflect market, counterparty and liquidity aspects.

In Tables from 3 to 6 we provide a summary description of the single components involved in the definition of the distress indicator. In particular, for each sub-market we show the average (quarterly or monthly) of daily prices observations for respectively: the logarithm of the ratio between the maximum and the minimum, the dispersion, the number of deals and the average notional traded amount. These statistics allow us to depict how sub-markets have evolved over time, thus suggesting the emergence of some common pattern that might have affected the overall behaviour as well as the presence of specific features that characterize certain sub-markets.

Descriptive statistics provide some insights on sub-markets' behaviours during the sample period. As regarding the (ln) max/min deviations, for some sub-markets (1, 2, 3, 4) the first part of 2014 coincides with lower mean values, while in the recent period they reach wider deviations. Conversely, other sub-markets (5, 6, 8, 9) show flattening or even declining trends during the reference period. These patterns are on average confirmed when we consider the estimates for dispersions. In addition, one might be interested in disentangle whether sub-markets with common contractual terms share similar trends. For instance, the absence of clearing agreements (sub-markets 7, 8, 9) seems to slightly affect the overall picture, since pairs of sub-markets (e.g. 5-8 and 6-9), which have the same maturity and the same swap frequencies legs but present different clearing agreements, exhibit close estimates. Furthermore, as we expect for sub-markets with high volumes of transactions, those with cleared conditions (from 4 to 6) show a smaller price dispersion than the respective un-cleared sub-markets (from 7 to 9). Finally, even for sub-markets with different swap leg frequencies but with the same maturity, we can observe that the price dispersion are quite close in some cases, e.g. sub-market 2 and the parallel sub-market 5 in the second part of the sample period. Moreover, it may be the case that a sub-market has a high max-min deviation but low dispersion (e.g. sub-market 4).

Table 3: Deviation between Maximum and Minimum. The value in a certain cell stands for the natural logarithm of the ratio between the maximum and the minimum of deals' prices for the corresponding sub-market and period. Values are averaged among daily observations, separately for each sub-market. Column names refer to the period (monthly or quarterly) considered to calculate mean values.

Sub-mkt	SEP 2013	IV Q 2013	I Q 2014	II Q 2014	III Q 2014	IV Q 2014	I Q 2015	APR 2015
1	0.35	0.14	0.10	0.15	0.37	0.37	0.49	0.37
2	0.29	0.24	0.19	0.58	0.71	0.65	0.54	0.55
3	0.05	0.04	0.04	0.19	0.23	0.23	0.20	0.27
4	0.55	0.42	0.54	0.48	0.75	0.86	0.81	0.64
5	1.06	1.06	1.08	1.00	0.83	0.81	0.70	0.76
6	0.43	0.41	0.33	0.34	0.30	0.31	0.42	0.36
7	0.54	0.51	0.80	0.57	0.72	0.66	0.79	0.30
8	1.10	0.98	1.09	0.99	0.83	0.79	0.68	0.62
9	0.53	0.44	0.39	0.33	0.36	0.34	0.33	0.40

Table 4: Price Dispersion. The value in a certain cell stands for the standard deviation of the prices for the contracts belonging to the corresponding sub-market and period. Values are averaged among daily observations, separately for each sub-market. Column names refer to the period (monthly or quarterly) considered to calculate mean values.

Sub-mkt	SEP 2013	IV Q 2013	I Q 2014	II Q 2014	III Q 2014	IV Q 2014	I Q 2015	APR 2015
1	0.34	0.11	0.07	0.04	0.11	0.10	0.15	0.12
2	0.29	0.26	0.24	0.30	0.32	0.29	0.21	0.21
3	0.20	0.10	0.09	0.25	0.28	0.25	0.17	0.20
4	0.06	0.04	0.04	0.05	0.09	0.09	0.11	0.08
5	0.43	0.40	0.46	0.43	0.36	0.32	0.25	0.24
6	0.33	0.37	0.34	0.32	0.28	0.25	0.19	0.19
7	0.15	0.16	0.20	0.13	0.17	0.17	0.22	0.10
8	0.53	0.49	0.51	0.48	0.41	0.36	0.31	0.28
9	0.44	0.41	0.37	0.34	0.31	0.27	0.24	0.34

The last two measures point out a picture characterised by increasing trends in both the number of deals and the average notional traded amounts, although estimates for the last period seem to indicate a re-

Table 5: Number of Traded Deals. The value in a certain cell stands for the number of deals corresponding to that sub-market and period. Values are averaged among daily observations, separately for each sub-market. Column names refer to the period (monthly or quarterly) considered to calculate mean values.

Sub-mkt	SEP 2013	IV Q 2013	I Q 2014	II Q 2014	III Q 2014	IV Q 2014	I Q 2015	APR 2015
1	2.29	1.73	1.48	4.20	5.74	7.45	5.13	4.35
2	3.09	3.02	2.66	24.23	51.01	46.22	35.94	31.58
3	1.80	1.96	2.64	25.29	46.41	47.21	46.97	37.61
4	24.71	28.55	39.15	33.18	40.83	57.19	50.27	48.36
5	183.16	252.00	292.24	257.18	294.99	332.27	309.83	244.12
6	195.48	241.54	228.13	223.52	272.65	311.14	320.49	251.65
7	6.41	5.25	6.34	5.17	7.24	9.59	8.47	4.48
8	25.41	21.04	25.61	24.43	33.17	40.68	22.50	16.50
9	30.15	25.04	21.62	18.39	35.75	44.10	21.27	13.60

Table 6: Notional Traded Amounts. The value in a certain cell stands for the average notional traded amount (in million of US Dollar) corresponding to that sub-market and period. Values are averaged among daily observations, separately for each sub-market. Column names refer to the period (monthly or quarterly) considered to calculate mean values.

Sub-mkt	SEP 2013	IV Q 2013	I Q 2014	II Q 2014	III Q 2014	IV Q 2014	I Q 2015	APR 2015
1	20.00	65.10	62.70	129.40	164.19	167.57	175.55	169.81
2	30.38	32.43	43.83	76.00	98.28	102.73	98.68	104.40
3	30.88	31.95	39.39	37.31	41.92	46.80	44.95	40.85
4	153.77	166.49	152.82	157.93	144.77	161.41	186.18	178.76
5	77.18	95.51	94.38	93.21	91.00	93.86	89.74	87.40
6	43.00	49.11	46.08	45.13	43.53	45.36	42.82	42.54
7	84.48	142.72	138.34	148.59	136.43	129.29	153.99	205.69
8	55.93	67.25	83.19	78.58	72.23	81.10	83.98	98.13
9	36.21	40.05	50.21	42.99	38.39	41.08	43.28	46.95

newed decrease in the number of transactions. In some cases (e.g. sub-markets 5-6) although the average cardinalities are similar, the average notional traded amounts are quite different. Sub-markets' heterogeneous dynamics across different dimensions suggest to consider a comprehensive set of measures instead of a single one to disentangle and characterize the level of distress present in a certain sub-market. Therefore, the overall picture provided by these figures indicate that a reasonable indicator of sub-market's conditions should rely on a combination of these measures.

For these reasons, we propose the following indicator of distress:

$$I_{i,t} = \ln\left(\frac{max_{i,t}}{min_{i,t}}\right) \times \frac{\sigma_{i,t}}{(Avgvolume_{i,t} \times Num_{i,t})} \tag{2.1}$$

where i and t are the indexes for the sub-markets and the weekly observations, respectively. Symbols *max* and *min* denote the maximum and the minimum of the weekly prices for each sub-market *i* at time *t*, respectively. Quantity $\ln\left(\frac{max_{i,t}}{min_{i,t}}\right)$ is lower-bounded and increases when the deviation between the *max* and the *min* becomes larger. The symbol σ stands for the standard deviation of the prices: its impact on the indicator of distress is positive, as greater volatility might be associated with distressed market conditions. Conversely, Num (i.e. the number of deals) has a negative effect since it is assumed that more traded deals imply that it is easier to find a counterparty, thus limiting liquidity risk. Lastly, the use of mean volumes ($Avgvolume$) indicates the average notional traded value of the deals and it is introduced for liquidity purposes. Therefore, we decide to consider explicitly each driver (i.e. the deviation max/min, the dispersion, the cardinality and the average traded amount) in the formula for the sake of clarity, although we are aware that there are some redundant issues related to the use of Num in the estimates of both dispersion and average volumes. Although market and liquidity risk drivers play an important role in the assessment of sub-market conditions, we prefer to provide a combined indicator of distress that aggregates a more comprehensive mix of effects. In addition, the choice to rely on these building block components reflects the idea that the deviation between the max and the min is a rough measure of liquidity conditions since it simply represents a couple of extreme points while ignores the stream of prices in the middle. Therefore, we decide to correct this estimate by introducing the price dispersion to mimic the effective distribution of the prices. Then, we further adjust this indicator by adding other two components in order to take into account the presence/lack of a sufficient number of deals (and/or average notional traded amount) and differentiate (*ceteris paribus* $\ln\left(\frac{max}{min}\right)$ and σ) between cases where the market is characterized by few deals (and/or with low average notional traded amount) and cases where we observe more deals and/or higher average notional traded amount[12].

We believe that is reasonable to rely on this simple indicator that captures in a qualitative way (increasing or decreasing indications) the impacts of the different distress factors, thus focusing on the discussion of the preliminary empirical results. Below, we show in Table 7 the average (monthly or quarterly) of the weekly observations of the indicator of distress as defined above to present how sub-markets' distresses have evolved over time. These estimates reflect the joint contributions of the single measures introduced above. In order to assess the level of distress within a sub-market one should in principle observe the magnitude of this measure, since by construction higher values correspond to deteriorated market conditions under our assumptions.

Table 7: Indicator of Distress. The value in a certain cell stands for the indicator of distress (values multiplied by 10^9) corresponding to that sub-market and period. Values are averaged among weekly observations, separately for each sub-market. Column names refer to the period (monthly or quarterly) considered to calculate mean values.

Sub-mkt	SEP 2013	IV Q 2013	I Q 2014	II Q 2014	III Q 2014	IV Q 2014	I Q 2015	APR 2015
1	12.21	1.63	0.44	0.24	0.16	0.11	0.17	0.10
2	9.61	1.97	1.88	0.28	0.05	0.05	0.12	0.04
3	0.09	0.32	0.05	0.28	0.04	0.03	0.02	0.04
4	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01
5	0.04	0.03	0.03	0.02	0.01	0.01	0.01	0.01
6	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
7	0.29	0.51	0.42	0.22	0.31	0.22	0.30	0.05
8	0.46	0.62	0.32	0.34	0.18	0.13	0.16	0.15
9	0.24	0.23	0.15	0.25	0.11	0.06	0.12	0.28

[12]In addition, one could argue that formula (1) can be improved by generalizing it with some parameters to be calibrated in some optimal way, such as:

$$I_{i,t}\left(\alpha,\beta,\gamma,\delta\right)=\ln\left(\frac{max_{i,t}}{min_{i,t}}\right)^{\alpha}\times\frac{\sigma_{i,}^{\beta}}{\left(Avgvolume_{i,t}^{\gamma}\times Num_{i,t}^{\delta}\right)}. \tag{2.2}$$

With regard to Table 7, we observe some stylized relevant facts[13]. The distress indicator by its construction does not have a practical or physical meaning although it allows some qualitative insights by looking at the *ranking* between the different markets. Hence it is worth noting that sub-markets from 4 to 6 (which involve *bank-to-bank* most liquid sub-markets) show a very low distress level. If we analyse the other sub-markets (from 7 to 9) with the same swap leg frequencies, it seems that the un-cleared ones (usually deals between *bank-to-enterprise*) exhibit more risky figures (this is mainly due to the lack of liquidity and/or large max-min range).

Basically, this indicator allows us to capture in a formal and intuitive way the causal forces that could move the sub-markets towards a distressed state. To switch from an useful but still descriptive representation into the investigation of how sub-markets are reciprocal influenced, we analyse how pairs of sub-markets are jointly dependent, i.e. how sub-markets' distresses co-move. Therefore, we attempt to study the dependence structure of the co-movements by computing, for each sub-market, the following increments of the indicator of distress:

$$X_{i,t+1} = \frac{I_{i,t+1} - I_{i,t}}{I_{i,t}} \tag{2.3}$$

for $i = 1, ..., S$ and $t = 0, ..., T-1$. Hence, a sub-market that exhibits positive increments implies that it experiences deteriorated conditions which became more serious if these variations become larger. For these reasons, our analysis is focused more on the right tail of the distributions of the increments, which corresponds to distressed market conditions.

[13]Estimates for September 2013 might even reflect the *backload* process of the deals. For instance, in EU the EMIR regulation was practically applied from February, 2014. At that time also the deals already alive were uploaded by a massive *backload* process. Hence we can doubt about the quality of the oldest data. In fact from the effective trade repository feed running process the distress indicators become lower and more stable. As a further remark, let us note that the *VIX* popular index, i.e. the volatility index of the S&P index level, did not reach at the end of 2014 any abnormal levels. In September 2013 the average level was 14.65%, just 50 bps higher than the average level of 2014, 14.14%.

2.4 Methodology

"Distress" is an extreme event, which can be seen as an upper tail event related to the process that describes the movement of the sub-market's status. In particular, we aim at providing, for each pair of sub-markets, a *Joint Probability of Distress* (JPoD)[14], that is the joint probability that both sub-markets simultaneously exhibit increments of the distress indicator above a certain threshold. This approach is similar to the one in Segoviano and Goodhart (2009), where the indicators known as Banking Stability Measures are presented. In fact, our sub-markets concept replaces their portfolio of financial institutions and, similarly to their study, we provide a distress interdependence structure which is able to capture not only linear correlations but also nonlinear distress dependencies among the players in the system.

To compute the joint probabilities of distress, we split the analysis into three parts. Firstly, once sub-markets have been set up, we study the form of association between each pair of sub-markets and how strong this relationship is. We exploit the family of Archimedean bivariate copulas (more precisely: *Clayton*, *Frank* and *Gumbel* copulas). The general theory of copulas[15] states that a joint distribution of some random variables can be decomposed into a function (called copula), that describes the interdependence structure among the considered variables, and their marginal distributions. The reason to narrow the choice of copula functions among Archimedean ones lies in the fact that we want the possible dependencies to be comparable. The Archimedean family provides, through a unique parameter (i.e. θ), a proxy for the dependence degree between the two sub-markets. Secondly, after having identified the dependence structures for each pair of sub-markets, we produce a ranking based on the Kendall's tau correlation coefficients. Finally, we compute joint probabilities of distress at different marginal threshold levels.

[14] The meaning of joint distress of couples of sub-markets, as well as the related terminology, are referred to concepts introduced in this paragraph and in Section 2.3. Hereinafter, any reference to existing expressions has to be considered contextualized to our work.

[15] As a reference to the copula theory, we rely on the well-known results provided in Sklar (1959), Nelsen (2006) and Trivedi and Zimmer (2007).

Note that in this work we exploit bivariate copulas to study dependencies among pairs of sub-markets. This choice is due to the small number of available sub-markets. Obviously, generalizations to multidimensional structures are possible and in Appendix A.2 we briefly discuss the case with copula dimension equal to 3.

In the following three subsections, we illustrate the technical details of the three steps of our study: the identification of the copula function for each possible pair of sub-markets, the global ranking classification based on Kendall's tau, and the computation of the JPoD for pairs of sub-markets (see Appendix A.1 for details). According to the latter probabilities, a final ranking classification of pairs of sub-markets is provided too.

2.4.1 The Preliminary Copula-based Procedure

Given S sub-markets and, for each sub-market, T time-observations of the random variable of interest X (described above by formula (2.3)), we can represent data by means of a real-valued matrix $\mathbf{X}$ of dimension $S \times T$,

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1t} & \cdots & x_{1T} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{it} & \cdots & x_{iT} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{S1} & \cdots & x_{St} & \cdots & x_{ST} \end{bmatrix} = \begin{bmatrix} x_{1\cdot} \\ \vdots \\ x_{i\cdot} \\ \vdots \\ x_{S\cdot} \end{bmatrix}$$

where x_{it} represents the value of the observation t for the sub-market i and $x_{i\cdot}$ is the row-vector that contains all the values related to sub-market i.

The preliminary procedure we propose takes as input this matrix and returns for each pair of sub-markets the most appropriate Archimedean copula and the corresponding parameter θ:

1. The procedure derives the marginal distribution for each sub-market i by finding the empirical cumulative distribution function $\widehat{F}_i$ based on the corresponding T-dimensional row $x_{i\cdot}$. For each sub-market i, we are assuming the values $x_{i1}, \ldots, x_{iT}$ as i.i.d. realizations drawn from the same univariate distribution.

2. Given a pair of different sub-markets, say (i, j), for each copula type (Cl =Clayton, Fr =Frank, Gu =Gumbel), the procedure computes the maximum value of the copula loglikelihood and the corresponding estimated value of the dependence parameter. Formally, it maximizes the function defined as $\theta \mapsto \ell_{(i,j),type}(\theta) = \sum_{t=1}^{T} \ln c_{type}\left(\widehat{F}_i(x_{it}), \widehat{F}_j(x_{jt}); \theta\right) + \sum_{t=1}^{T} \left(\ln f_i(x_{it}) + \ln f_j(x_{jt})\right)$,

   (note that the second term does not depend on θ, nor on $type$) where $c_{type}(u_1, u_2; \theta)$ denotes the parametric expression of the density for the chosen copula ($type \in \{Cl, Fr, Gu\}$), and it records the values $\ell^*_{(i,j),type}$ and $\theta^*_{(i,j),type}$ such that

   $$\ell^*_{(i,j),type} = \ell_{(i,j),type}(\theta^*_{(i,j),type}) = \max_{\theta \in \Theta} \ell_{(i,j),type}(\theta) .$$

   Note that we are considering pairs $\{(x_{it}, x_{jt}) : t = 1, \ldots, T\}$ as T i.i.d. realizations drawn from the same bidimensional distribution.

3. For each possible pair (i, j) of different sub-markets, the procedure finds $\ell^*_{(i,j)}$, $\theta^*_{(i,j)}$, and $type^*_{(i,j)}$ such that

   $$\ell^*_{(i,j)} = \max_{type \in \{Cl, Gu, Fr\}} \ell^*_{(i,j),type} \tag{2.4}$$

   and $\theta^*_{(i,j)}$ and $type^*_{(i,j)}$ are the corresponding estimated parameter and the corresponding selected copula-type, respectively.

   The maximization (2.4) corresponds to select the copula that provides the best fit according to both AIC and SIC criteria[16]. Indeed, we have the best fit at the lowest value of the quantity

[16] For further references see Mahfoud (2012).

$$\begin{aligned}
AIC &= -2 \times (loglikelihood) + 2 \times (n.\,parameters) \\
&= -2 \times (loglikelihood) + 2 \\
SIC &= -2 \times (loglikelihood) + \ln(n.\,observations) \times (n.\,parameters) \\
&= -2 \times (loglikelihood) + \ln(T),
\end{aligned}$$

respectively, and so at the highest value of the loglikelihood.

2.4.2 The Correlation Ranking

For each possible pair (i, j) of different sub-markets, the first step of the procedure selects the copula function, i.e. the type of the copula ($type^*_{(i,j)}$) and the respective parameter ($\theta^*_{(i,j)}$). The goal of the second step is to produce a classification of the most dependent pairs of sub-markets. One possible choice to measure the strength of the dependence between two sub-markets relies on their parameter $\theta^*_{(i,j)}$; for the Archimedean family of copulas, indeed, it gives a measure of dependence between (i, j). However, the theta parameter is related to the functional form of the copula and so values of the theta parameter for different copula functions are not directly comparable. Therefore, in order to circumscribe this issue, we use the value of Kendall's tau[16] for each pair as the criterion for the ranking. Denoting by $\tau^*_{(i,j)}$ the value of the Kendall's tau coefficient as a function of $\theta^*_{(i,j)}$ (see Appendix A.1.2), our procedure considers each possible pair of different sub-markets and splits the final ranking of the pairs of sub-markets into two groups: the pairs with positive Kendall's tau dependence coefficient (i.e. $\tau^*_{(i,j)} \geq 0$) and the ones with negative dependence coefficient (i.e. $\tau^*_{(i,j)} < 0$). Finally, it returns: a decreasing ranking of the pairs of sub-markets based on $\tau^*_{(i,j)}$ for the first group, and, for the second group, an increasing ranking of the pairs based on the (negative) value of $\tau^*_{(i,j)}$. (Note that negative dependence parameter is possible only for $type \in \{Cl, Fr\}$.)

Together with the two classes of rankings (positive and negative), the procedure returns for each pair (i, j):

- $type^*_{(i,j)}$ (based on the following code: $1 = Fr$, $2 = Gu$, $3 = Cl$);

- the value of the difference $diff\_theta_{(i,j)} = (\theta^*_{(i,j)} - \theta_{type^*_{(i,j)},ind})$ where $\theta_{type^*_{(i,j)},ind}$ is the value for the chosen copula $type^*_{(i,j)}$ corresponding to the independence case[17],

- the estimated value for the Kendall's tau $\tau^*_{(i,j)}$ as a function of the theta parameter for the selected copula;

- the empirical value $e\_\tau^*_{(i,j)}$ of the Kendall's tau.

2.4.3 Joint Probability of Distress (JPoD)

Once checked the appropriateness of the selected copula model, the analysis is carried out on computing, for each pair of sub-markets, the joint probability that both of them simultaneously exhibit increments of the distress indicator above some given threshold, i.e. the joint probability of distress. We calculate this probability at different marginal threshold levels. In our context, we prefer to avoid the selection of a specific "distress threshold" and we decide to analyse sub-markets' joint behaviours in the right tail at different marginal levels. More precisely, if we denote by X_i and X_j the increments of the distress indicator (defined in Section 2.3) for sub-markets i and j respectively, then, for each pair (x, y) of real numbers, we have:

$$\begin{aligned} P(X_i > x, X_j > y) &= 1 - P(X_i \leq x \, or \, X_j \leq y) \\ &= 1 - F_i(x) - F_j(y) + P(X_i \leq x, X_j \leq y) \\ &= 1 - F_i(x) - F_j(y) + F(x, y) \\ &= 1 - F_i(x) - F_j(y) + C(F_i(x), F_j(y)), \end{aligned}$$

[17] The parameters which correspond to the independence case are: 0 (asymptotic value) for the Frank and the Clayton copulas, 1 for the Gumbel copula.

where F_i and F_j are the marginal cumulative distribution functions, F is the joint cumulative distribution function of the pair (i, j) and the last equality is due the Sklar's Theorem (see Appendix A.1). Consequently, we define our Joint Probability of Distress (JPoD) as:

$$JPoD_{(i,j)} = 1 - u_i - u_j + C_{type^*_{(i,j)}}(u_i, u_j; \theta^*_{(i,j)})$$

where $u_i, u_j \in [0, 1]$ are the levels for the marginal cumulative distribution functions F_i, F_j, typically chosen equal to 90%, 95% and 99%.

2.5 Results

In the next Tables, sub-markets are referred to the previous classification (see Table 1). Copula types are expressed as: 1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*). In addition, *"diff_theta"* refers to the theta parameter returned once a copula type is chosen minus the value of the theta parameter for the independence case for this type of copula. Moreover, the "e" before the parameter refers to the empirical estimates (when no type of copula is imposed but estimates are computed on raw data). Our perimeter is composed by 25 pairs of sub-markets which show positive estimated Kendall's tau correlations and 11 sub-markets with negative values. For the sake of clarity, we consider only about the first half of the rankings, i.e. the first 10 and 5 pairs for positive and negative Kendall's tau, respectively. Therefore, we focus on those pairs of sub-markets which show estimates more distant from the independent case.

Positive Kendall's tau estimates show a very interesting behaviour if we focus on the pairs of sub-markets in the first positions of the ranking. In particular, we can rewrite the 9 sub-markets by an integer triple $M_j,\ j = 1...9$ as follows:

$$M_j = (f_j, t_j, c_j) \tag{2.5}$$

where:

- f = frequency, 0 = 3m-3m, 1 = 6m-3m
- t = tenor range, 0 = short, 1 = medium, 2 = long
- c = clearing, 0 = cleared, 1 = un-cleared.

Hence, sub-markets span a very simple discrete space where we could define between each pair a *Manhattan*-like distance, such as:

$$d(M_i, M_j) \equiv |f_j - f_i| + |t_j - t_i| + |c_j - c_i| \,. \tag{2.6}$$

Given this simple framework we can study more in depth the positive estimates of Kendall's tau values. In particular, we can observe that all the first four (with respect to the Kendall's tau metrics) pairs have the minimum distance between their components, i.e. $d(M_i, M_j) = 1$. This an appealing empirical fact, since despite several issues related to the difficulty to identify sub-markets, such as the pioneering work on TRs data and the new distress indicator definition, these preliminary outcomes are near to the practical intuition. In addition, as shown by Table 9 even negative estimates can occur. This is the case of pairs of sub-markets with quite different maturities and clearing conditions. Hence, sub-markets with different features are more prone to show opposite co-movements, while as expected similarities in financial contractual terms are more likely to determine positive and high co-movements.

Finally, as shown in Tables 8 and 9 for both positive and negative Kendall's tau rankings, we observe that there is a very high correlation among the empirical Kendall's tau ($e\_\tau^*_{(i,j)}$), which is calculated on the two vectors not processed through the copula procedure we decided to apply, and the Kendall's tau ($\tau^*_{(i,j)}$), which we obtain according to the type of copula chosen for the pair of sub-markets (i, j) and its estimated parameter $\theta^*_{(i,j)}$. In particular, this correlation is equal to 0.991 for the positive ranking table and to 0.955 for the negative one. This suggests that the copula selection procedure provides a good fit.

Table 8: Distress Indicator: *Positive* Kendall's Tau. Ranking of reciprocal co-movements. Ranking is shown in a descending ordering based on positive Kendall's tau. Columns *I Sub-mkt* and *II Sub-mkt* stand for the pair of sub-markets selected by our procedure. Column *Copula* refers to the chosen copula type (1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*)). Column *diff_theta* stands for the theta parameter returned once a copula type is chosen minus the value of the theta parameter for the independence case. Finally, the empirical Kendall's tau is shown in the last column.

Ranking	I Sub-mkt	II Sub-mkt	Copula	diff_theta	Kendall's tau	e_Kendall's tau
1	5	6	2	0.48	0.32	0.32
2	5	8	3	0.65	0.24	0.24
3	2	3	3	0.48	0.19	0.20
4	4	5	3	0.45	0.18	0.16
5	1	6	3	0.36	0.15	0.14
6	8	9	1	1.09	0.12	0.12
7	2	7	3	0.25	0.11	0.09
8	1	8	2	0.12	0.11	0.11
9	4	8	1	0.99	0.11	0.11
10	1	5	3	0.23	0.10	0.08

Table 9: Distress Indicator: *Negative* Kendall's Tau. Ranking of reciprocal co-movements. Ranking is shown in a ascending ordering based on negative Kendall's tau. Columns *I Sub-mkt* and *II Sub-mkt* stand for the pair of sub-markets selected by our procedure. Column *Copula* refers to the chosen copula type (1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*)). Column *diff_theta* stands for the theta parameter returned once a copula type is chosen minus the value of the theta parameter for the independence case. Finally, the empirical Kendall's tau is shown in the last column.

Ranking	I Sub-mkt	II Sub-mkt	Copula	diff_theta	Kendall's tau	e_Kendall's tau
1	5	7	1	-2.23	-0.24	-0.23
2	3	8	3	-0.33	-0.20	-0.24
3	1	3	1	-1.52	-0.16	-0.17
4	7	8	3	-0.22	-0.12	-0.06
5	6	7	1	-0.98	-0.11	-0.10

Moreover, in the last three columns of Table 10 we report the JPoD at marginal levels for F_i and F_j both equal to 90%, 95% and 99%, respectively. In particular, since we attempt to study the joint probability of distress, we focus on those pairs of sub-markets which exhibit positive Kendall's tau values (see Tables 8). In addition, the selection of these thresholds is intended to provide some insights of these relationships at the tail of the distribution corresponding to deteriorated market conditions. Preliminary results suggest that for the first two pairs the JPoD assumes quite relevant values, while for the other positions estimates are almost comparable. For instance, the first pair of sub-markets in the ranking position is $(i, j) = (5, 6)$, meaning that this pair of sub-markets has the most correlated increases in terms of percentage of the distress indicators (I_i, I_j) at 90%, 95%, 99% levels for the marginal cumulative distribution functions. These sub-markets shares the same swap legs (6m3m), the same cleared conditions (cleared contracts in both sub-markets), but have different maturities (Medium vs Long). In addition, they represent the two most active segments in the IRS market, as reported in Table 2. Therefore, it seems that the two most important sub-markets in terms of number of deals and traded notional amounts are also highly co-dependent. At first glance, if we look at the following positions we can observe slightly different rankings compared to those shown in Tables 8. However, it is worth highlighting that JPoD and Kendall's tau rankings are coherent once we focus on a certain type of copula, i.e. given the same copula we observe that the ordering for pairs of sub-markets is the same for both types of rankings. Finally, we briefly analyse the second position in Table 10, that is $(1, 8)$. This pair of sub-markets has different swap legs (3m3m vs 6m3m), different cleared conditions (cleared vs un-cleared), and different maturities (Short vs Medium). Still, they share a high probability of joint distress, thus supporting the need of further investigation on the features that can impact on the reciprocal influence between sub-markets apparently very distant. Hence, similarities between financial and contractual terms seem to be responsible for stronger co-dependences in many cases, although the emergence of high values for JPoD in correspondence of quite different sub-markets (pair 1,8) suggests the presence of other reasons than contractual terms. In particular, this underlines the need to identify the key players operating in these OTC IRS markets, since their roles may influence co-movements between apparently different sub-markets.

Table 10: JPoD at Different Marginal Threshold Levels. Ranking based on JPoD estimates for different levels of thresholds. Ranking is shown in a descending ordering based on JPoDs. Columns *I Sub-mkt* and *II Sub-mkt* refer to the pairs of sub-markets selected by our procedure. Column *Copula* stands for the chosen type of copula (1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*)). Column *Tau Ranking* refers to the ranking based on the procedure exploited to select the type of copula for each pair of sub-markets. Finally, the last three columns refer to the JPoD associated to different levels of the threshold (90%, 95% and 99%).

JPoD Ranking	Tau Ranking	I Sub-mkt	II Sub-mkt	Copula	JPoD 90%	JPoD 95%	JPoD 99%
1	1	5	6	2	4.4963%	2.1259%	0.4059%
2	8	1	8	2	2.2563%	0.9288%	0.1540%
3	2	5	8	3	1.5457%	0.3985%	0.0164%
4	6	8	9	1	1.4875%	0.3901%	0.0163%
5	9	4	8	1	1.4381%	0.3755%	0.0156%
6	3	2	3	3	1.4142%	0.3619%	0.0148%
7	4	4	5	3	1.3843%	0.3537%	0.0144%
8	5	1	6	3	1.3089%	0.3330%	0.0135%
9	7	2	7	3	1.2176%	0.3082%	0.0124%
10	10	1	5	3	1.2032%	0.3043%	0.0123%

2.6 Conclusions and Future Research

The financial crisis of the last decade motivated a growing literature on how to model and to predict the financial distress. Some concepts such as the *systemic risk*, the *contagion* effect and the *cascade* defaults received an increasing attention. Nevertheless, a "new normal" for the risk management field has not yet been established. If we consider the financial system as a whole, several challenging aspects have yet to be solved, such as: the huge number of risk factors and financial products, their dependence structures, the lack of complete and granular data about the financial system, the quality of available data, the measures to be used to capture and predict markets' co-movements.

To partially address these issues, we exploited and combined in an innovative way some new ingredients, namely the OTC derivatives data provided by trade repositories along with the JPoD approach recently introduced by the International Monetary Fund. To the best of our knowledge, this is the first attempt that exploits micro-founded data from trade repositories to study co-dependence phenomena between financial sub-markets. In particular, we focused on the interest rate derivatives as a significant fraction of the OTC market and we defined a distress indicator by combining four different distress drivers, such as liquidity, average traded volumes, volatility and bid-ask proxies. Hence, we attempted to study by this framework the *distress dependencies* of some OTC sub-markets that we built according to contractual and financial features. By analysing both the descriptive results and the Joint Probabilities of Distress, the proposed technique seems to be quite promising for assessing markets' co-movements in a financial stability perspective. Despite its complex work flow, preliminary results are close to the practical intuition. Similarities between financial and contractual terms seem to be responsible for stronger co-dependences, although high values for JPoD even in correspondence of quite dissimilar sub-markets suggest the presence of other drivers which need to be investigated in future research (such as the role of key market players active across different sub-markets that cannot be identified in our dataset).

We also remark the need for a sharp distress definition to calibrate a more general distress indicator formula which can be applied for *back-testing* procedures, i.e. to assess its prediction properties. Furthermore, in the future other asset classes (equity, credit, forex, etc..) could be exploited to implement a financial "classical" top-down sub-markets segmentation. Finally, some deeper knowledge about the TRs effective internal data quality is required.

Appendix A

A.1 Copula Methodology

Copula functions provides a mathematical tool for the modelization of the multivariate stochastic dependence structure. In particular, copulas take into account various kinds of stochastic dependence structures among actors, without any assumption on the one-dimensional marginal distributions. The concept of copula was introduced during the forties and the fifties with Sklar (1959), but the evidence of a growing interest in this kind of functions in statistics started only in the nineties (see Hoeffding (1994) and Nelsen (2006)). Copulas are functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distributions. The advantage of the copula functions and the reason why they are used in the dependence modeling is related to the Sklar's theorem (see Sklar (1959)). It essentially states that every multivariate cumulative distribution function can be rewritten in terms of the margins, i.e. the marginal cumulative distribution functions, and a copula. More precisely, we have the following definition and results.

A d-dimensional copula $C(\mathbf{u}) = C(u_1, ..., u_d)$ is a function defined on $[0,1]^d$ with values in $[0,1]$, which satisfies the following three properties:

1. $C(1, ...1, u_i, 1, ..., 1) = u_i$ for every $i \in \{1, ..., d\}$ and $u_i \in [0,1]$;

2. if $u_i = 0$ for at least one i, then $C(u_1, \ldots, u_d) = 0$;

3. for every $(a_1, ..., a_d), (b_1, ...b_d) \in [0,1]^d$ with $a_i \leq b_i$ for all i,

$$\sum_{j_1=1}^{2} ... \sum_{j_d=1}^{2} (-1)^{j_1+...+j_d} C(u_{1,j_1}, ..., u_{d,j_d}) \geq 0$$

where, for each i, $u_{i,1} = a_i$ and $u_{i,2} = b_i$.

Let F be a multivariate cumulative distribution function with margins $F_1, ..., F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for every $x_1, ..., x_d \in \overline{\mathbb{R}} = [-\infty, +\infty]$, we have

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{A.1}$$

If the margins $F_1, \ldots, F_d$ are all continuous, then C is unique; otherwise C is uniquely determined on $F_1(\overline{\mathbb{R}}) \times \cdots \times F_d(\overline{\mathbb{R}})$.

Conversely, if C is a copula and $F_1, \ldots, F_d$ are cumulative distribution functions, then F defined by (A.1) is a multivariate cumulative distribution function with margins $F_1, \ldots, F_d$.

In the case when f and $f_1, \ldots, f_d$ are the marginal probability density functions associated to F and $F_1, \ldots, F_d$, respectively, the copula density c satisfies

$$f(x_1, \ldots, x_d) = c\big(F_1(x_1), \ldots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i).$$

There are different families of copula functions that capture different aspects of the dependence structure: positive and negative dependence, symmetry, heaviness of tail dependence and so on. In our work, we limit ourselves to the principal copula functions of the Archimedean family (namely, Gumbel, Clayton and Frank copulas), which describe, through a unique parameter θ, situations with different degrees of dependence.

For more details on copula theory, we refer to the various excellent monographs existing in literature, such as Joe (1997), Nelsen (2006) and Trivedi and Zimmer (2007).

A.1.1 Archimedean family of copulas

Here we just recall, in the bivariate case, the principal copula functions belonging to the Archimedean family that we employ in our analysis (Huynh et al. 2014).

- **Frank copula**:

$$C_{Fr}(u_1, u_2; \theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{\exp(-\theta) - 1} \right),$$

$$\theta \in \Theta = (-\infty, +\infty) \setminus \{0\}.$$

The parameter θ tunes the degree of the dependence. The limiting cases $\theta \to \theta_{Fr,ind} = 0$ correspond to independence.

- **Gumbel copula**:

$$C_{Gu}(u_1, u_2; \theta) = \exp\left\{ - \left[(-\ln u_1)^\theta + (-\ln u_2)^\theta \right]^{\frac{1}{\theta}} \right\},$$

$$\theta \in \Theta = [1, +\infty).$$

The parameter θ tunes the degree of the dependence. In particular, the value $\theta = \theta_{Gu,ind} = 1$ corresponds to independence (indeed, we get $C^{Gu}(\mathbf{u}; 1) = \prod_{i=1}^{d} u_i$).

- **Clayton copula**:

$$C_{Cl}(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-\frac{1}{\theta}},$$

$$\theta \in \Theta = [-1, +\infty) \setminus \{0\}.$$

The parameter θ controls the degree of the dependence. The limiting case $\theta \to \theta_{Cl,ind} = 0$ corresponds to independence.

A.1.2 Kendall's tau

Consider two random variables X, Y with continuous marginals F_1, F_2 and joint cumulative distribution function F [1]. The Kendall's tau correlation coefficient is defined as:

$$\tau(X,Y) = P\left\{(X_1 - X_2)(Y_1 - Y_2) > 0\right\} - P\left\{(X_1 - X_2)(Y_1 - Y_2) < 0\right\}$$

where (X_1, Y_1) and (X_2, Y_2) are two independent pairs of random variables from the joint distribution F. It can be written in terms of the copula function as follows:

$$\tau(X,Y) = 4\int_0^1\int_0^1 C(u_1,u_2)\,dC(u_1,u_2) - 1.$$

In particular, for the Archimedean copulas, the Kendall's tau can be expressed as a function of the dependence parameter θ:

$$\tau(X,Y) = \begin{cases} 1 + 4\theta^{-1}[\theta^{-1}\int_0^\theta t/(e^t-1)\,dt - 1] & \text{Frank} \\ 1 - \theta^{-1} & \text{Gumbel} \\ \theta/(\theta+2) & \text{Clayton.} \end{cases} \tag{A.2}$$

[1] For further details, see Trivedi and Zimmer (2007).

A.2 Multivariate Example

In the paper we focused on copula dimension equal to 2. Obviously, generalizations to higher dimensions are feasible although it is worthwhile to recall that in our case we are dealing with only 9 sub-markets. Below we briefly report the trivariate case and we present similar estimates to the ones shown in Table 8. The algorithm still gives the possibility to choose among three different copulas (Frank, Gumbel and Clayton). We estimated all possible triple results, namely 84 positions (84 are indeed the combinations of 9 sub-market), although for the sake of conciseness we report only the first ten positions. Both tables show the value of the parameter estimated by our procedure minus the parameter for the independence case (*diff_theta*). In particular, Table 12 is ordered by decreasing maximum likelihood, while Table 11 is obtained by ordering the triples by decreasing *diff_theta*.

Terns which appear in the first ten positions of both tables should be pointed out in order to make a financial analysis of the sub-markets, since those should represent the most trustful results being evidenced from two different ordering criteria. We observe that the triples (4,5,8), (5,6,8) and (1,5,6) appear in both Table 11 and 12. The first two pairs share similar contractual terms, i.e. they refer to swaps frequency legs equal to (6m3m) and present a short Manhattan-like distance (it is 4 in both cases, computing by summing distances among each couple in the tern). Conversely, tern (1,5,6) shows quite different features and presents a higher distance (it is 6). Thus, even in the trivariate case dissimilarity among contractual and financial terms can imply high co-movements. Below we analyse the tables more in detail providing a comparison with the sub-markets which appeared to be co-dependent in the bivariate case as shown in Table 8.

Ranking in Table 11 is based on *diff_theta*. Estimates are coherent to those exhibited in Table 8: in the first positions we observe combinations of pairs (5,6), (5,8), (2,3) and (8,9), i.e. pairs of sub-markets that are strongly co-dependent in the bivariate case are more likely to influence co-movements also in the trivariate case. Hence, relevant relationships among sub-markets seem to emerge regardless the dimension of the copula. In addition, we can also revise results analogies with the ones provided by Table 10, which gives the output of JPoD (not implemented in the trivariate case): we find again that sub-markets (5,6), (5,8), (2,3) and (8,9) characterize the top positions in the ranking. Finally, in Table 12 we show how sub-markets are ranked based on the maximization of the log-likelihood. Even in this case (similarly as discussed above in Section 2.5), results are coherent among Tables 11 and 12 once we consider the estimates within the chosen type of copula. Overall, this is a further evidence of the robustness of our procedure; however, increasing too much the copula dimension may result meaningless when having few sub-markets.

Table 11: Ranking based on Diff_Theta (Top 10 Positions). Ranking of co-movements when copula dimension is equal to 3. Ranking is shown in a descending ordering based on positive *diff_theta*. Columns *I Sub-mkt*, *II Sub-mkt* and *III Sub-mkt* stand for the triple of sub-markets selected by our procedure. Column *Copula* refers to the chosen copula type (1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*)). Column *diff_theta* stands for the theta parameter returned once a copula type is chosen minus the value of the theta parameter for the independence case. Column *Ranking ML* refers to the ranking based on maximization of loglikelihood. Finally, we report for each combination the respective max loglikelihood.

Ranking ML	I Sub-mkt	II Sub-mkt	III Sub-mkt	Loglikelihood	Copula	diff_theta
18	2	5	6	3.81	1	1.11
19	3	5	6	3.60	1	1.07
21	2	3	6	3.21	1	1.02
24	4	8	9	2.79	1	0.91
36	1	8	9	1.65	1	0.70
1	4	5	8	10.54	3	0.48
2	5	6	8	9.69	3	0.44
57	3	5	9	0.60	1	0.42
61	2	4	7	0.47	1	0.37
3	1	5	6	7.70	3	0.37

Table 12: Ranking based on Max Log-Likelihood (Top 10 Positions). Ranking of co-movements when copula dimension is equal to 3. Ranking is shown in a descending ordering based on the maximized loglikelihood. Columns *I Sub-mkt*, *II Sub-mkt* and *III Sub-mkt* stand for the triple of submarkets selected by our procedure. Column *Copula* refers to the chosen copula type (1 (*Frank*), 2 (*Gumbel*), and 3 (*Clayton*)). Column *diff_theta* stands for the theta parameter returned once a copula type is chosen minus the value of the theta parameter for the independence case.

Ranking	I Sub-mkt	II Sub-mkt	III Sub-mkt	Loglikelihood	Copula	diff_theta
1	4	5	8	10.54	3	0.48
2	5	6	8	9.69	3	0.44
3	1	5	6	7.70	3	0.37
4	1	5	7	7.70	1	0.37
5	4	5	6	7.23	3	0.36
6	4	5	7	7.23	1	0.36
7	5	8	9	7.21	3	0.37
8	6	7	8	7.21	1	0.37
9	6	7	9	7.21	1	0.37
10	1	5	8	6.08	3	0.33

Chapter 3

The Accounting Network: how Financial Institutions react to Systemic Crisis

3.1 Introduction

Network Theory has been used to establish how contagion, through a variety of channels (mutual exposures, social networks of board members, moral hazard from permissive regulations, financial instruments like swaps and derivatives, etc.), triggered the outbreak of the 2007-08 crisis. Scholars suggest that financial systems may affect positively economic development and its stability (Beck 2009; Beck 2011; Levine 2005), although they may represent a source of distress which leads to bank failures and currency crises, or greater contraction for those sectors that depend more on external finance during banking crisis (Dell'Ariccia et al. 2008; Reinhart and Rogoff 2009). As a response to the recent financial turmoil, banking sector has been affected by a substantial reorganization (BIS 2014). For instance, as highlighted by the European Central Bank for the Euro area *the main findings reflect the efforts by banks to rationalize banking businesses, pressure to cut costs, and the deleveraging process that the banking sector has been undergoing since the start of the financial crisis* *in 2008* (ECB 2013). This implies that market pressure and regulatory amendments induce banks to reduce their levels of debt, through cost containment and stricter capital requirements. In addition, a gradual improvement in banks' capital positions aims to enhance the capacity of the system to absorb shocks arising from financial and economic distresses. This limits the risk of spillover effects from the financial sector to the real economy and put the financial system in a better condition to reap the benefits of economic recovery. In particular, as the financial boom turned to a bust, banks' stability deteriorated abruptly and the economy entered a *balance sheet recession*, which depressed spending levels through a reduction in consumption by households and investments by firms. Therefore, although at an uneven pace across regulations, the need to strengthen fundamentals has influenced banking sector, and differences in banks' portfolio allocations, financial performances, and capitalizations can be interpreted as the combined results of policy decisions and sectoral responses to changes in the regulatory framework (see e.g. Allen, Gu, et al. (2012), Diamond and Rajan 2009).

This paper relates to the literature on banking development and performance evaluation during the recent crisis (see e.g. Adrian and Shin 2008, Berger and Bouwman 2013 and Brunnermeier 2009). We consider a dataset of almost one thousand worldwide banks retrieved from *Bloomberg*, focusing on financial statements spanning from 2001 to 2013. We introduce a network based on similarities between banks' financial statement compositions (hereinafter *Accounting Network*). Due to data limitations, the reference sample is restricted to banks for which a continuum and stable set of variables is available for the entire period. The introduction of a methodology (*Quality Ratios*) to measure banks' data coverage aims to prevent that missing values for some variables or lack of annual financial statements for some banks affect the overall picture. We then exploit the maximum amount of available information from financial statements without further reducing the set of variables through an arbitrary selection of financial statements fields. This choice aims to avoid any selection bias. Moreover, total assets (as a proxy for size) for each bank is applied to normalize banks' financial statements measures to prevent the emergence of "size effects" as the sizes of institutions are spanning for various orders of magnitude.

The introduction of *Accounting Networks* establishes a bridge between the external perspective arising from market data and the internal one based on banking activities indicators. We study how *Accounting Networks* can be exploited to provide a description of the banking system during the crisis. This part sheds light on whether banks under different regulatory frameworks and diversification degrees have reacted to the crisis by strengthening their business peculiarities or by converging towards similar practices (Beltratti and Stultz 2012, Demirgüc-Kunt and Huizinga 2010). We rely on the assumption that market data alone, despite highly representative of investors' perception of the banking sector, might be dis-informative during periods of distressed market conditions. This, in turn, stimulates a broader exploitation of the information on banking activities, thus pointing to a more comprehensive investigation which takes into account also the internal perspective arising from financial statements data. In addition, the use of accounting data allows a partition of business activities where banks are involved in, providing therefore an approximation of the state of the system related to several potential channels through which financial distress might propagate. This is appealing also for regulators, since authorities are interested in a wide set of economic indicators in order to prevent the systemic relevance of financial institutions and they introduce specific requirements and constraints which affect directly financial statements measures. For these reasons, we believe that enriching the debate on financial stability by means of the *Accounting Networks* might give new clues about the resilience of the banking system.

Another important result is the possibility of getting a neutral partition of banks in "network communities" that result from the analysis of the network through community detection algorithms like the *Louvain* modularity maximization (Blondel et al. 2008). Results in Subsection 3.3.1 indicate that regional communities evolve in time and the crisis has a clear role in weakening geographically determined structures. Furthermore, we focus on proxies for leverage, size and performance to verify if these variables have played a key role among the set of economic measures usually applied to classify banks (see e.g. Blundell-Wignall and Roulet 2013, Huizinga and Laeven 2012). Hence, Subsection 3.3.2 aims to answer the question whether the collapse of financial markets has weakened these relationships, limiting therefore the power of traditional indicators to identify groups of homogeneous banks. Correlation diagrams applied to show how network variables are related to economic measures suggest a turning point in correspondence of the outbreak of the crisis, which influenced the role of proxies for leverage, size or performance to group similar banks. This preliminary results motivated Subsection 3.3.3, where by means of Principal Components Analysis we investigate which economic features are more likely to characterise communities' heterogeneity before, during and after the collapse of 2007-08.

The remaining part of the work discusses open issues and future lines of research (Section 3.4), such as open questions on how to improve the building of the *Accounting Networks*. In particular, the effectiveness of this approach can be enhanced by means of a careful variables selection based on the best financial practices applied in the evaluation of the financial statements structures. In addition, a more accurate normalization of the variables and caring about national regulations may increase the usefulness of the methodology. Furthermore, matrix filtering techniques and missing data reconstruction for financial statements information can enhance the extraction of meaningful clusters. Then, more advanced and focused tools could be conceived to analyse banks evolution towards similar business configurations or, alternatively, their divergent patterns as a response to changing market conditions.

3.2 Methods

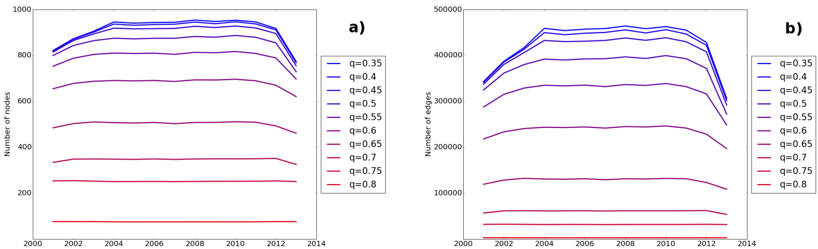
3.2.1 Dataset Preparation

The dataset we analysed covers the set of banks provided by *Bloomberg* which were active (i.e. with traded instruments) at the end of the first quarter of 2014. Although quarterly information is available, we prefer to focus on annual balance sheets and income statements for accounting standard reasons, as countries have different obligations in terms of the provision of quarterly financial statements and this can lead to a mismatch and a poor variables coverage. Data are collected during the reference period from 2001 to end of 2013.

As regards financial statements data, we select a large set of variables among those available in *Bloomberg* and related to the current regulatory framework (BCBS 2011a). We rely on the existing literature for the selection process, providing a neutral approach which does not weight differently the financial statements items. We focus the analysis on proxies for banking business models (see e.g. Altunbas et al. 2011, Calomiris and Nissim 2014). In particular, balance sheet data provide a year-by-year picture of stock variables in terms of assets and liabilities for different instruments and maturities, while income statement data describe annual economic performances by partitioning profits and losses according to banking activities ranging for instance from interests to fees. Since national regulations allow firms to fix a different end of fiscal year, we extend the "end of year" definition and the relative financial statements according to a window in the range between three months before and after the end of the solar year. Solving overlapping issues in variables definitions, as well as the base currency choice, constitute the first step in the data pre-processing procedure. Firstly we discard total and sub-total measures (as they are redundant measures), and secondly we choose US dollars as currency base, thus facilitating banks comparisons (for the list of variables see Appendix 14).

Working with financial statements data often leads to limitations in data coverage and completeness. Therefore, the starting point of our analysis is represented by the selection of a stable set of banks in terms of data availability during the sample period. In particular, banks might change the composition of their financial statements or they might be excluded by the *Bloomberg* provider due to several reasons, such as for instance a new regulation or a change in the bank's economic activities. This, in turn, might cause *missing values* for some variables or lack of financial statements for several banks in certain years. In order to limit the impact of these issues on our findings, we define a methodology to measure the coverage of available variables for each bank in the reference period. We refer to the *Quality Ratios (QRs)* as the proportion of available and usable variables V_{OK} over the maximum of all possible ones V_{ALL} in the sample period: $QR = V_{OK}/V_{ALL}$. The tuning of this indicator, combined with two more filters on the frequency of financial reporting, provides a stable set of banks identified by their QRs. The two additional criteria are: a minimum number of financial statements of ten out of thirteen possible fiscal years and a maximum gap period between two consecutive annual reports equal to seven hundred days. Once selected those banks that report almost continuously their financial statements, we study them according to their respective QR.

Figure 3: Quality Ratios Distribution. These pictures show the number of nodes (on the left) and edges (on the right) along the sample period for different QR values.



Individual QRs, as empirically computed on the entire perimeter, lie in the range between 0.3 (low accuracy/coverage) and 0.8 (high accuracy/coverage). Interestingly, many measures calculated on the sets of banks obtained by fixing the QR do not seem to be significantly affected by its choice (except, as expected, for high QRs, where the size of the sample reduces significantly). With greater values of the QR parameter we have less available banks to be considered, since only few of them have a large set of variables present in many of their financial statements. As estimates are stable in a reasonable QR range, in this work we decide to use the set arising from the case of QR = 0.5 that, even if arbitrary, still represents a good compromise between the accuracy of the estimate and the size of the sample (see Figure 3).

3.2.2 Accounting Networks

For every year a vector of financial statement variables is assigned to each bank and used to compute the cosine similarities between pairs of banks/nodes. Here the intuition is that the most similar banks (as from their financial statements) must stay closer in the network and form a cluster. Then, the measure "cosine similarity" is transformed into a metrics. The definition is the following: we compute the cosine of the angle between each pair of vectors with the dot product and then we apply the simple transformation $w_{i,j} = 1 - \sqrt{1 - C_{i,j}^2}$, where $w_{i,j} \in [0, 1]$ and $C_{i,j}$ is the cosine similarity between *i* and *j*. In network terms $w_{i,j}$ is the weight. This transformation (see Dongen and Enright 2012 for an introduction to similarity measures and relative metrics) is used to move from the cosine similarities defined in the space [-1, +1] to weights in the interval [0, +1]. Actually, in our networks cosine similarities range mainly between 0 and +1, therefore with this transformation the more two nodes are similar the larger is the weight, while a weight of 0 is assigned to dissimilar pair of nodes having basically orthogonal financial statements' vectors.

In addition, before the computation of the metrics, we need to take care of banks' size distribution, as it spans over several orders of magnitude. To avoid a bias toward large institutions, for each bank we divide all variables in its vector by the respective total assets in such a way that the attributes of the vector refer to economic and financial *ratios*. This operation ensures that clusters will be formed by banks with similar business activities regardless their size levels.

An important methodological choice of our study is the "neutral" approach used for the selection of the variables within the financial statements. A part from removing related and redundant measures (total and subtotal items), we used all the available information applying the same weight to each variable in the vectors. This agnostic approach is in line with the goal of the paper, i.e. introducing the concept of *Accounting Network*, although we are aware that practitioners can give a different importance to each variable of the financial statement to assess the true business model. In our perspective we expect that the relevant information will emerge in a bottom up process, as a spontaneous feature selection carried by our methodology.

Finally, we introduce a confidence level (95%) during the link formation. By using a Montecarlo sampling test, if the cosine similarity is statistically significant with 95% of confidence we retain the link otherwise we discard it. As a result of this filtering procedure, we observe that networks tend to be very dense and almost complete. The most of the information is carried by the weights of the links and less by the simple topology (degrees and other structural features).

3.2.3 Community Detection

A classical method to investigate the structure of a network is the search of communities, i.e. regions of the network with larger *internal* links density. Intuitively, these regions are formed by clusters of nodes with higher degrees or, for weighted networks, with larger strengths. Several methods have been proposed to find network communities without imposing a priori the number of communities but letting them emerging from the network itself. Among others we cite the optimization of the modularity that is a measure of how much the link structure differs from the random network where links are assigned with uniform probability and internal communities are not present (a part from fluctuation). For weighted networks, the modularity is defined by the following formula:

$$Q_w = \frac{1}{2W} \cdot \sum_{ij} \left(w_{ij} - \frac{s_i s_j}{2W} \right) \delta(c_i, c_j) \tag{3.1}$$

where $s_i = \sum_j w_{ij}$ and $s_j = \sum_i w_{ij}$ are the strengths (sum of weights) of the nodes i and j respectively, W is half the sum of the weights in the network and the function $\delta(c_i, c_j)$ is equal to 1 if (i, j) belong to the same community or 0 if they are members of different communities. The maximum modularity value is 1 (an ideal case for which the communities are isolated) and can also take negative values. The 0 value coincides with a single partition that will correspond to the whole graph. A negative value means that there is no particular advantage in separating the nodes in those particular clusters and so there is not community structure whatsoever.

To study the presence of communities it is often necessary to prune the network cutting the links if their weight is below a certain threshold. In our case we intend to consider only links formed by nodes having a large similarity/weight w of their financial statement vectors. The pruning procedure can be guided by the use of the tools related to the community detection methodology (Fortunato 2010). In particular, working with the modularity optimization function (Newman and Girvan 2004), with the *Louvain* technique (Blondel et al. 2008), it is possible to look at the *significance* associated to the threshold (as in Traag et al. 2013), where the modularity is introduced as a parameter to check for the best resolved community structure. We use this parameter to help finding a reasonable pruning threshold range of values for the networks. A rule of thumb in this process is indeed avoiding network fragmentation, i.e. keeping the graph connected while removing not significant links. We made extensive tests computing quality/significance of the partitions using different pruning thresholds (i.e. removing links having a low weight), determining a range of weights thresholds ($0.35 < w_{i,j} < 0.6$) that helps to prune the original networks to an optimal level (see Appendix B.1). In this interval, communities are stable and the interpretation of each region can be seen as a result of financial statement similarities across banks in different countries.

3.2.4 Network Measures vs. Economic Indicators

Comparisons among network measures and economic indicators are provided to describe the correlation between nodes' network topology and economic behavior. We study these features by means of extensive linear correlation tests (Pearson correlation) for the overall set of banks for each year and we verify the significance of the estimates by means of parametric tests. These estimates are based on the filtered networks, which are themselves based on the significance and the quality of the community detection configuration. This analysis shows how nodes' network properties (e.g. *Strength* or *Average Clustering Coefficient*) are associated to basic economic indicators (e.g. *Return on Assets*, *Total Assets* and *Total Debts to Total Assets*), thus showing whether nodes' topological properties are positively or negatively related to certain economic features and how these relationships have weakened or reinforced during the crisis.

The average clustering coefficient is the mean of the measure of the local tendency of the nodes to form small regions of fully connected nodes. The in- (out-)strength of a node is the weighted sum of its in- (out-)coming links. Return on assets (ROA) is the net income over total assets and is a measure of bank performance. Total debts to total assets is an indicator of the leverage of the bank and it is computed as the ratio between total debts and its size (measured by total assets).

3.2.5 Principal Components Analysis

Once communities are identified, we attempt to describe which financial statement variables are more likely to characterise these clusters. In order to facilitate comparability, we focus on those indicators more popular within the set of variables utilised to compute the cosine similarities (i.e. those indicators appearing with higher frequencies in the entire dataset). In fact the inclusion of very poorly represented measures across different banks would have made the comparisons less effective with potential biases related to e.g. different regulations frameworks or geographical memberships. Hence, since we are interested in disentangling potential similarities/peculiarities across different communities, we prefer to rely on common and well-diffused measures of banking activities among those present in banks' financial statements. In addition, we enrich this set by means of indicators such as ratios (e.g. *Return on cap* and *Total debts to total assets*) and aggregated measures (e.g. *Total assets*). Community detection identifies four main clusters, whose constituents are more numerous and stable in time. For the sake of conciseness, Subsection 3.3.3 will focus mainly on these communities. In particular, for each year we describe by means of Principal Components Analysis (PCA) which economic features are more (less) able to contribute to the explained variability of communities' members.

PCA is a multivariate technique that analyses observations described by several inter-correlated variables. It extracts the important information from the data and expresses it as a set of new orthogonal variables (principal components). In our exercise, since measures present different ranges of dispersion (e.g. by construction some ratios are bounded) we rely on a scaled version of PCA; finally, we consider only principal components with eigenvalues greater than 1 (in almost all cases they correspond to the first 3 components). Then, we compute the proportion of the variance of each original economic measure that can be explained by the selected principal components. This, in turn, leads to a ranking of the original economic measures in terms of their ability to describe a certain community's variability. In particular, since we are interested in how the onset of financial crisis has affected the banking system, we split this analysis in three periods: from 2001 to 2006 (before the crisis), from 2007 to 2009 (the onset of the crisis), and from 2010 to 2013 (after the breakdown of the markets). For each period we decided to characterise each community by the top and the bottom three measures, thus analysing how these ranks have evolved over time and across communities.

3.3 Results

This Section shows how *Accounting Networks* represent a complementary approach to traditional financial networks[1] for the study of banking system. While financial networks reflect the view from the market, related to for instance the pairwise correlations of banks' stock prices, *Accounting Networks* capture the effects of business decisions on financial statements measures and on business models of different institutions. An "ideal" in-

[1]There is a huge and growing literature on financial networks built on market data, where links are computed for instance according to stock prices correlations. Payment system represents another well-studied channel for contagion and cascade effects, while the exploitation of derivatives (e.g. CDS) data is a cornerstone of many works on systemic risk.

vestigation of the financial system would involve also a detailed analysis of the money flows among institutions, which determine the so called "mutual exposures" (an important contagion channel). Unfortunately, these high granular and detailed data are usually not available. However, financial statements provide an aggregated view of mutual exposures and obligations for different maturities and types of financial instruments. This is an important point in favour of *Accounting Networks* as they report summarised information for e.g. phenomena occurring with different time scales and contractual terms, as opposite to financial networks that rely only on homogeneous (daily or intraday) market data.

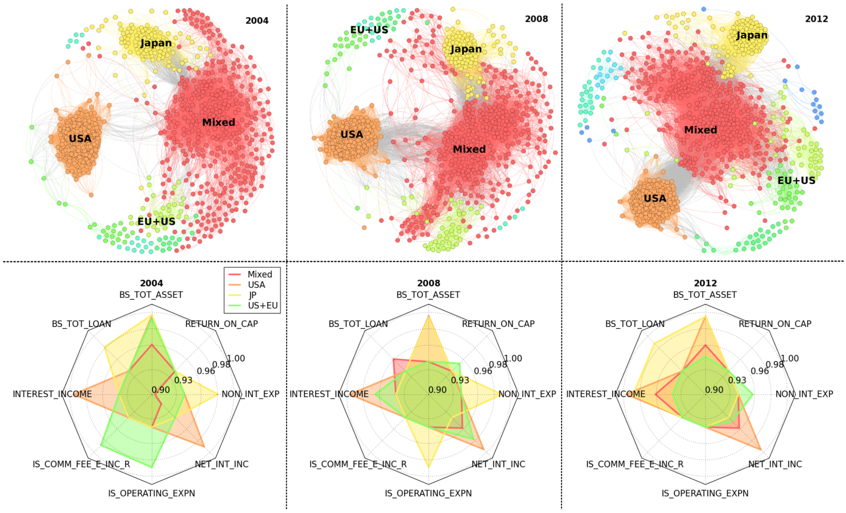
3.3.1 Community Detection Results

In this Subsection we focus our attention on the bottom up clusterization of the network from the application of the community detection algorithm and on the presence of geographical structures arising when we label each bank with its country of origin. We describe whether banks belonging to different countries (as a proxy for different regulations and/or level playing fields) have shown the tendency to be part of separate or, alternatively, common clusters and we verify, by analysing communities' evolution over time, whether the crisis influenced these configurations. In particular, our community detection analysis on *Accounting Networks* shows these main results (see Appendix B.3 for details on communities membership).

It exhibits the presence over time of a clear community representing US banks and another one composed by Japanese banks, although for both regions there is also an additional smaller second group quite persistent in time. By contrast, it is not possible to identify a single and an unambiguous European community, since institutions belonging to European countries seem to be more likely to form national or sub-regional communities or to be included in a vast and geographically heterogeneous cluster (hereinafter the *Mixed* community). Asian banks are fragmented in several sub-regions where, in particular, the Arab and the Indian-Pakistan groups emerge. Therefore, the detection of communities within *Accounting Networks* reveals the presence of two homogeneous clusters corresponding to US and Japanese banks surrounded by a more diversified cloud of institutions belonging to different countries; remarkably, European banks are not able to clusterise together in a single and distinct community, while it persists over time a certain level of separation based also on national borders. Hence, an interesting contribution of the paper points to the presence of a large and geographically heterogeneous community, which can be related to the fact that the globally established regulatory framework might have indeed accelerated the tendency of banking activities of different countries to converge into more uniform banking practices (see e.g. BCBS 2011a; BIS 2014; ECB 2013). This is shown for instance in Figure 4 where we also observe that the outbreak of financial markets contributed to make the *Mixed* community more cohesive; furthermore, although still representing separate communities, both *US* and *JP* clusters result topologically closer to the *Mixed* community after the breakdown of 2007-08, thus supporting the interpretation of a gradual convergence of different areas into more similar banking behaviours. Moreover, the application of the community detection on *Accounting Networks* allows to identify even small communities, such as those related to African or Scandinavian banks. This represents a quite promising aspect of the methodology, since it ensures the detection of local reliable communities although the approach taken so far is eminently agnostic.

It is not simple to explain the reasons behind the emergence and evolution of these communities, however it is possible to advance some intuitions based on the impact of globally recognized accounting standards (for a time line see e.g. FASB 2016), the establishment of supranational supervisory and regulatory authorities, and on the role of the harmonization process of banking practices which have been implemented through, for instance, the various Basel regulations (BCBS 2011a). These contributions point to a common level playing field, which might have facilitated the emergence of a large and geographically heterogeneous community and its increasing topological proximity to both *US* and *JP* clusters. However, latter communities highlight the persistence of regional

Figure 4: Community Detection Results. In the upper plots we show the Community Structures which arise from the application of the *Louvain* algorithm (see subsection 3.2.3). In the radar plots we exhibit the most important financial statements components by the PCA analysis (see subsection 3.2.5). We consider years 2004, 2008 and 2012 corresponding to the sub-periods Pre-Crisis, Crisis and Post-Crisis. Both representations indicate an increasing similarity in the system after the crisis: graphs show more cohesive communities which point to the *Mixed* cluster; radar plots exhibit communities' peculiarities which became more overlapped in 2012.



peculiarities. In Japan a deregulation process, known as the 'Japanese Big Bang', was formulated during the 1990s to transform the traditional bank-centered system into a market-centered financial system characterised by more transparent and liberalised financial markets (Hoshi and Kashyap 1999). In fact, peculiar features of Japanese banking sector were the over-reliance on intermediated bank lending, the absence of a sufficient corporate bond market and a marginal role for non-bank financial institutions, whose main consequences were an abundance of non-performing loans, excess in liquidity, scarce investments and low banks profitability (see e.g. Batten and Szilagyi 2003). Although this program was intended to cover the period 1996-2001, the goals have not been achieved yet and policy makers' continuing reform efforts to remove past practices by market participants confirm the slowing convergence of the Japanese regulatory framework to a capital-market based financial system (Aronson 2011). Thus, the presence of the *JP* community which gradually tends to the *Mixed* cluster is in line with evidences from the Japanese financial sector's reforms aimed to change its reliance on indirect finance into a system of direct finance related to capital markets. Furthermore, it is remarkable the presence of a *US* community quite stable over time, which seems to be progressively attracted by the *Mixed* cluster. The US financial system presents peculiar features compared to other geographical areas. It is characterised by a relatively greater role of capital market-based intermediation, a higher importance of the 'shadow banking system', and differences in the accounting standards (ECB 2013). The impact of non-bank financial intermediation relates to the use of originate-to-distribute lending models, which determine the direct issuance of asset-backed securities and the transfers of loans to government-sponsored enterprises (e.g. Fannie Mae and Freddie Mac), a phenomenon particularly relevant before the onset of financial crisis. Financial innovation played a key role in the early 2000s and the increasing use of securitisation explains the low percentage of loans to households on banks' balance sheets (ECB 2013). In addition, the 'shadow banking system' is highly dependent on the presence of finance companies, money market funds, hedge funds and investment funds, which influenced the growth of total assets in the US financial sector during the last decades (Shin 2012, Pozsar et al. 2012). The presence of a distinct community is probably also due to differences in accounting standards, which mainly involve the treatment of derivatives positions between the US Generally Accepted Accounting Principles (US GAAP) and the International Financial Reporting Standards (IFRSs). In particular, US GAAP allows to report the net value of derivative positions with the same counterparty under the presence of a single master agreement, thus impacting on the size representation of balance sheets items. However, in Figure 4 we observe that the *US* community (similarly to the *JP* community) is gradually approaching the *Mixed* community, and the consequences of the crisis seem to have enhanced this convergence. Among the possible several reasons, it is worth considering the impacts of the reform on the OTC derivatives market (embedded in the Dodd-Frank Act) and the Basel III new banking regulation, which may have facilitated similarities among US institutions and their peers in the *Mixed* cluster.

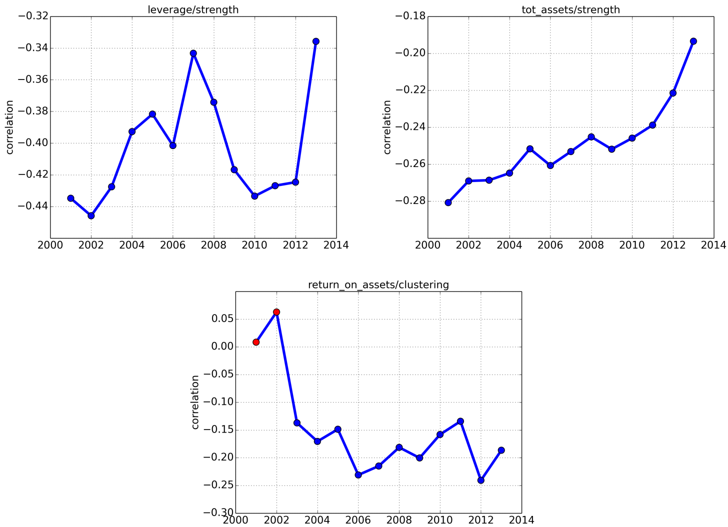
3.3.2 Relationships between Economic Indicators and Network Properties

In this Subsection we provide a preliminary investigation of the relationships between banks' economic indicators and their network properties. In order to characterise banks, we consider three common proxies for their classification: *Return on Assets* (for the *Performance*), *Total Assets* (for the *Size*) and *Total Debts to Total Assets* (for the *Leverage*). Then, comparisons are presented against two basic network measures: the *Strength* and the *Average Clustering Coefficient*. For each year from 2001 to 2013, we provide some insights for these relationships by estimating for the overall sample the correlations between banks' economic indicators and network measures. As explained in Subsection 3.2.4, in this exercise we consider networks filtered according to the quality/significance of the *Louvain* community detection algorithm, which helps us in the assessment of the significance of our results. Below, we show some examples to discuss how these relationships have evolved over time.

In particular, we investigate whether once the effects of the crisis have spread throughout the financial system, the capacity of traditional economic indicators (i.e. leverage, size, performance) to group banks resulted undermined. For instance, the onset of the crisis clearly affects the relationships between *Total Debts to Total Assets* and network properties. Although the correlation between *Strength* and *Total Debts to Total Assets* remains negative during the entire sample period, the breakdown of financial markets seems to further enhance this effect for subsequent years (Figure 5, plot on the top). Thus, this relationship suggests that, after the onset of the crisis, the use of leverage became on average more anti-correlated to the *Strength*. This implies that banks that are more dissimilar in terms of their financial statements (i.e. with lower values of *Strength*) are those that turned out to be less capitalised (i.e. with higher values of *Total Debts to Total Assets*). Furthermore, one might be interested in understanding the role played by the *Size* which represents a typical indicator utilised to classify banks. The correlation between *Strength* and *Total Assets* is negative and almost flat even after the collapse of 2007-08, while it shows an increasing trend in the recent period (Figure 5, plot on the middle). Hence, it seems that after the outbreak of the crisis the *Size* became less correlated to similarities among banks, as estimates pointing sharply to zero seem to suggest. We finally analyse the relationship between *Performance* and network properties (Figure 5, plot on the bottom). To determine the level of structure in the system we consider the *Average Clustering Coefficient* which mimics how the presence/absence of very connected groups of banks, that stand for the levels of "diversification" in the adoption of similar business models, might be related to economic results. Although poorly statistically significant in the early 2000s, correlations with *Return on Assets* exhibit a decreasing pattern before the onset of the crisis and then remain negative although slightly erratic. The negative relationship between *Average Clustering Coefficient* and *Return on Assets* seems to suggest that belonging to well connected areas of the network (nodes with higher clustering coefficients) do not foster economic performance.

These basic examples suggest that a clear investigation on the relationships between economic indicators and network properties may be not always conclusive and there might be some cases where estimates are poorly significant. Still, some remarkable effects arise from this investigation strategy and preliminary results point to a turning point in the correlations across the outbreak of the financial crisis. In particular, diagrams confirm that leverage is a useful indicator for differentiating banks, hence deviations to lower capitalizations are associated to increasing dissimilarities with the rest of the system and the impact of the crisis suggests a reinforcement of this relationship. By contrast, after the breakdown of 2007-08 it seems that size does not contribute too much on the similarity between banks, while it played a greater role before and during the crisis. Finally, the relationship between performance and the structure of the system is less clear and prevents straightforward conclusions.

Figure 5: Relationships between Economic Indicators and Topological Properties. In these plots we present the correlations between banks' Strength versus the Total Debts to Total Assets (Leverage) (plot at the top-left), Strength versus Total Assets (Size) (plot at the top-right) and Clustering Coefficient versus Return on Assets (Performance) (plot on the bottom part of the panel). The correlation is computed across the years 2001-13. Red points stand for not significant estimates at 5% level.



The study of the correlations between network properties and economic indicators reveals that the system has reacted against the crisis by updating the traditional drivers that can be used to reasonably classify banks. These preliminary findings are, therefore, intended to show that within an almost stable set of banks (thanks to the choice of the QR), the dynamics which drive the similarities of business models are actually influenced by the outbreak of financial markets. The identification of economic features potentially able to characterise specific portions of the system is addressed in the next Subsection.

3.3.3 PCA results

Community detection shows the presence of three large clusters (*Mixed*, *US*, and *JP*) and an additional quite stable and persistent but smaller community (mostly *US+EU* banks). In this Section we provide a way to describe how these communities can be represented in terms of economic features (see Figure 4). Given the multi-dimensionality of the set of measures utilised to build the networks, we adopt the Principal Components Analysis to identify those measures which contribute more (less) to the explained variance within each community. For the sake of simplicity, we propose the ranking of the top (bottom) three measures for each community during the following intervals: pre-crisis (2001-2006), crisis (2007-2009), and post-crisis (2010-2013). In particular, for each year we compute the contribution of the original measures to explained variance; then, we average within each sub-period and we determine the rankings based on the mean period values. Below, we name the community with a mixed geographical composition as *C0*, while we refer to the communities with a prevalence of *US*, *JP* and *EU+US* banks as *C1*, *C2* and *C3*, respectively.

This representation allows us to compare communities' features over time and across different groups. For instance, in Table 13 we observe that *Total Assets* and *Interest Income* are quite frequent among top measures contributors, while *Total Debts to Total Assets* is recurrent among measures in the bottom rankings. This is not surprising given banks heterogeneity in terms of the size (*Total Assets*) and the economic results (*Interest Income*) distributions, in contrast with the tight constraints on leverage (*Total Debts to Total Assets*) due to regulatory requirements. By focusing on the top rankings we notice that *C0* and *C1* have fairly stable top contributors, while communities *C2* and *C3* are more affected by the wave of financial turmoil. Furthermore, bottom rankings seem to be on average only slightly influenced by the choice of different sub-periods. In addition, differences between mean values among the set of top three and the set of bottom three contributors are quite stable over time with only few exceptions, while the middle part of the distribution of measures' contributions (not reported, available from authors upon request) is in general quite sparse. For these reasons, we prefer to focus on the top and the bottom rankings to describe communities' features.

One might be interested in how the outbreak of financial crisis have influenced these rankings. Top composition of *C1* is unaffected by the 2007-08 financial breakdown, while *C0* is only partially modified by the onset of the crisis (*Interest Income* is replaced by *Net Interest Income*). Conversely, *C2* presents a quite different configuration during the crisis sub-period when it exhibits a relevant role for expenses measures (i.e. Non Interest Expenses and Operating Expenses). Similarly, income statement measures become more relevant among top contributors also within the *C3* community. Interestingly, community *C0*, which is characterised by a mixed geographical composition, and the *US* community (*C1*) reach identical top contributors after the outbreak of 2007-08, while the *JP* community (*C2*), which shows the same top contributors as community *C0* in the first sub-period, seems to react differently during the crisis, although in the third sub-period it shows again top contributors similar to *C0* (and to *C1*). By contrast, community *C3* seems to present a peculiar pattern over time.

Therefore, the crisis sub-period coincides with remarkable differences in the top contributors, while the recent sub-period points to a renewed tendency to get similar contributors for a wider set of banks (*C0* and *C1*, and partially *C2*). This is in line with the above discussion on community detection results, where we highlighted a gradual proximity between clusters over time. Hence, these results suggest that heterogeneity within clusters is driven by similar economic measures after the crisis, although specific differences persist. This is the case for instance of loans, which are not present among top contributors in the *US* community while they are present in the top ranking of both the *Mixed* and the *JP* community (as expected according to the above discussion). PCA reveals that heterogeneity during the crisis sub-period have been heavily influenced by the presence of other economic dimensions than previous sub-period; in addition, it shows how after the crisis, these economic variables are more similar across communities; finally, peculiarities on top contributors confirm previous findings on geographical community detection.

Table 13: Characterization of Clusters in terms of Financial Statements Items. First Table shows the sets of top three contributors for each community, while the second Table shows the bottom three contributors. Values represent the contributions of original measures to the explained variances. Rankings refer to averaged values along each sub-period: 2001-06, 2007-09 and 2010-13. Community *C0* refers to the *Mixed* community, while *C1*, *C2* and *C3* stand for *US*, *JP*, and *EU+US* clusters, respectively.

Community	Top Measures 2001-06	Values 2001-06	Top Measures 2007-09	Values 2007-09	Top Measures 2010-13	Values 2010-13
C0	BS_TOT_ASSET	0.9695	BS_TOT_LOAN	0.9588	INTEREST_INCOME	0.9635
C0	BS_TOT_LOAN	0.9257	NET_INT_INC	0.9529	NET_INT_INC	0.9620
C0	INTEREST_INCOME	0.9228	BS_TOT_ASSET	0.9492	BS_TOT_ASSET	0.9537
C1	INTEREST_INCOME	0.9955	BS_TOT_ASSET	0.9964	INTEREST_INCOME	0.9968
C1	BS_TOT_ASSET	0.9951	INTEREST_INCOME	0.9957	NET_INT_INC	0.9954
C1	NET_INT_INC	0.9917	NET_INT_INC	0.9933	BS_TOT_ASSET	0.9953
C2	BS_TOT_ASSET	0.9935	BS_TOT_ASSET	0.9927	BS_TOT_ASSET	0.9943
C2	BS_TOT_LOAN	0.9854	NON_INT_EXP	0.9886	NON_INT_EXP	0.9877
C2	INTEREST_INCOME	0.9770	IS_OPERATING_EXPN	0.9883	INTEREST_INCOME	0.9876
C3	BS_TOT_ASSET	0.9817	INTEREST_INCOME	0.9678	NON_INT_EXP	0.9671
C3	IS_COMM_AND_FEE_EARN_INC_REO	0.9803	NON_INT_EXP	0.9670	NET_INT_INC	0.9621
C3	NON_INT_EXP	0.9800	IS_OPERATING_EXPN	0.9624	IS_OPERATING_EXPN	0.9564

Community	Bottom Measures 2001-06	Values 2001-06	Bottom Measures 2007-09	Values 2007-09	Bottom Measures 2010-13	Values 2010-13
C0	BS_LT_BORROW	0.7196	BS_LT_BORROW	0.6723	BS_LT_BORROW	0.7185
C0	BS_SH_CAP_AND_APIC	0.7131	BS_ST_BORROW	0.6511	TOT_DEBT_TO_TOT_ASSET	0.6659
C0	TOT_DEBT_TO_TOT_ASSET	0.4386	TOT_DEBT_TO_TOT_ASSET	0.5360	BS_ST_BORROW	0.6537
C1	RETURN_ON_ASSET	0.7006	BS_LT_INVEST	0.5799	BS_SH_CAP_AND_APIC	0.8151
C1	INTERBANKING_ASSETS	0.4987	INTERBANKING_ASSETS	0.5724	BS_LT_INVEST	0.7075
C1	TOT_DEBT_TO_TOT_ASSET	0.4941	TOT_DEBT_TO_TOT_ASSET	0.1366	TOT_DEBT_TO_TOT_ASSET	0.1762
C2	INTERBANKING_ASSETS	0.8878	RETURN_ON_CAP	0.7911	BS_ST_BORROW	0.8892
C2	TOT_DEBT_TO_TOT_ASSET	0.6980	BS_SH_CAP_AND_APIC	0.7245	TOT_DEBT_TO_TOT_ASSET	0.7574
C2	BS_SH_CAP_AND_APIC	0.5491	TOT_DEBT_TO_TOT_ASSET	0.5702	RETURN_ON_CAP	0.7498
C3	BS_SH_CAP_AND_APIC	0.8124	BS_CASH_NEAR_CASH_ITEM	0.8004	BS_CASH_NEAR_CASH_ITEM	0.7045
C3	RETURN_ON_ASSET	0.8007	BS_NON_PERFORM_ASSET	0.7707	IS_INT_EXPENSES	0.6626
C3	TOT_DEBT_TO_TOT_ASSET	0.6345	TOT_DEBT_TO_TOT_ASSET	0.5311	TOT_DEBT_TO_TOT_ASSET	0.6519

We also notice that the crisis suggests an increasing importance of income statement measures in terms of contribution to the explained variance within communities. The breakdown of financial markets affected banks' results and this justifies the high level of heterogeneity expressed by income statements indicators. This can be related also to the impact of the crisis on financial statement measures and, in particular, on the different ways banks update their balance sheet structures compared to the recognition of economic results as reported in the income statements items. Similar comparisons can involve also the bottom three measures, but for conciseness we omit this part.

3.4 Discussion

This paper describes banking system through similarities among banks' financial statements. Our main contribution is the introduction of a methodology to exploit balance sheets and income statements data to construct the *Accounting Networks*. We show some relationships between economic indicators and network properties, which might provide some new useful insights for banking classification practices, and we describe the emergence of geographical communities which reacted to the 2007-08 crisis converging to more similar banking practices. Depicting some effects of the recent financial crisis by using a simple framework is an encouraging sign for further extensions.

We rely on "neutral" and "naive" techniques to build the *Accounting Networks*. In particular, among common approaches usually applied to describe similarities concepts, we adopt one of the basic method, i.e. the cosine similarity. Future works can exploit more advanced methodologies. Moreover, our selection of variables utilised to compute cosine similarities assumes that each component has the same importance. This is quite a naive hypothesis, which could be enriched by the discrimination of economic indicators based on economic literature and/or practitioners perspectives. Finally, for accounting reasons we limit our study on annual financial statements, while a more detailed description of the system might easily involve the use of quarterly data. Despite these simplifying assumptions, our approach has the merit of introducing a novelty in the debate on banking networks, and we believe that future improvements in the directions outlined above will enforce *Accounting Networks*' ability to describe the evolution of banking systems.

Appendix B

B.1 Quality/Significance of the clustering for different pruning thresholds

Below, we report for each year the quality and the significance of the configuration arising from the application of the community detection algorithm on a pruned graph (where the thresholds are shown in the x-axis).

Figure 6: Quality/Significance for year 2001

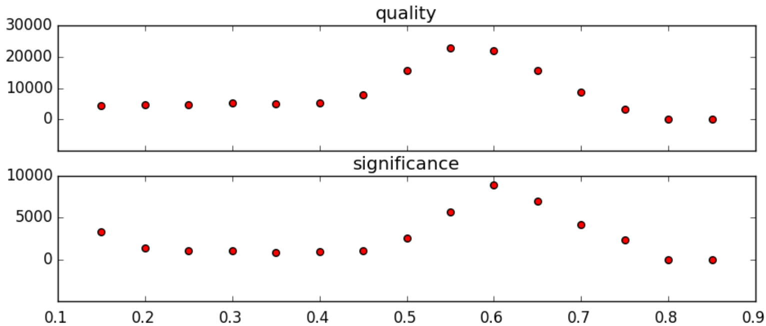


Figure 7: Quality/Significance for year 2002

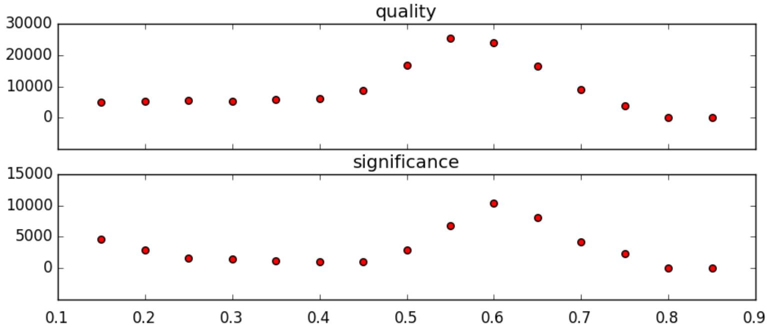


Figure 8: Quality/Significance for year 2003

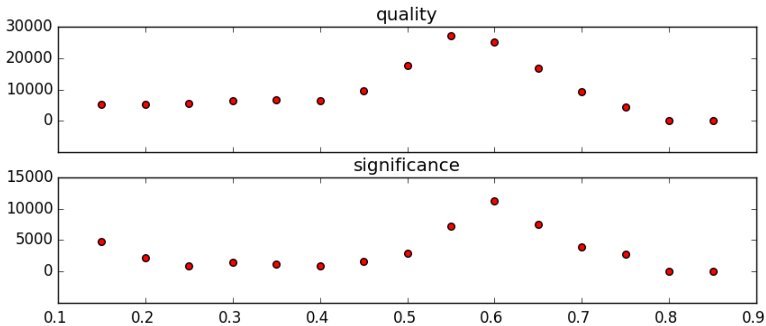


Figure 9: Quality/Significance for year 2004

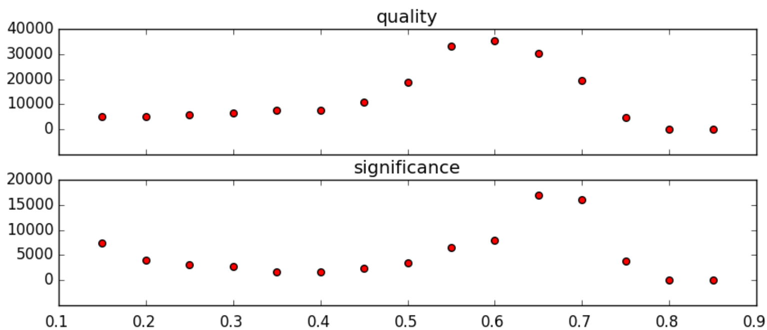


Figure 10: Quality/Significance for year 2005

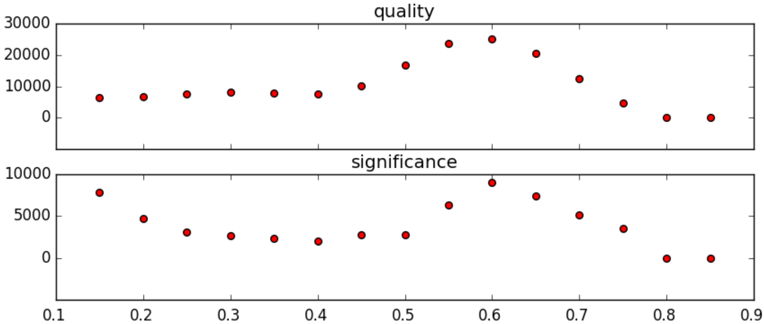


Figure 11: Quality/Significance for year 2006

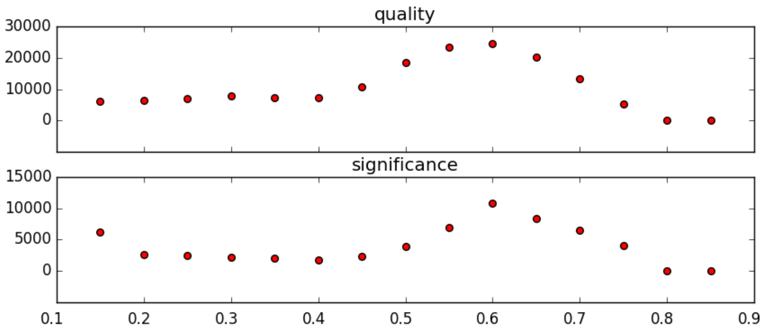


Figure 12: Quality/Significance for year 2007

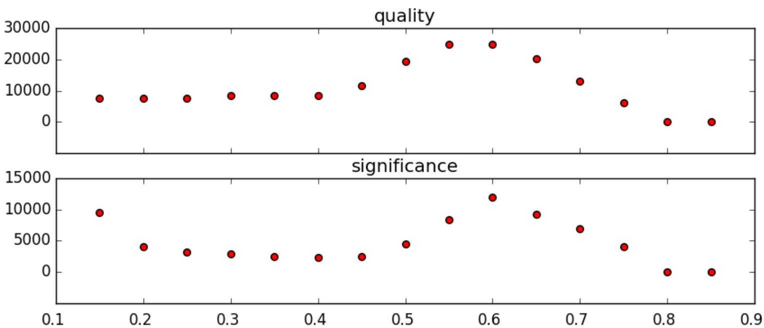


Figure 13: Quality/Significance for year 2008

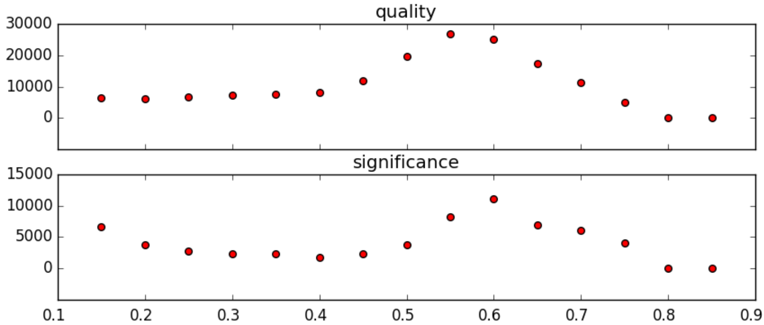


Figure 14: Quality/Significance for year 2009

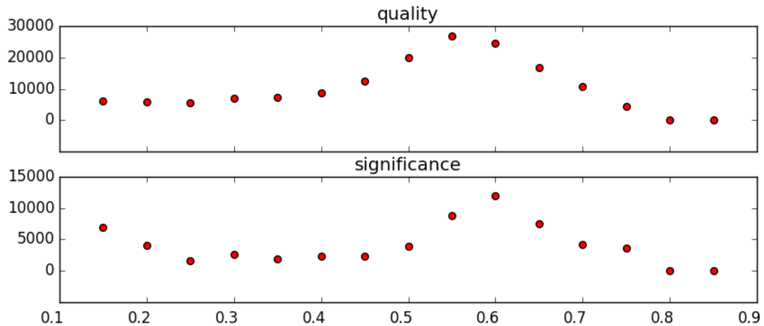


Figure 15: Quality/Significance for year 2010

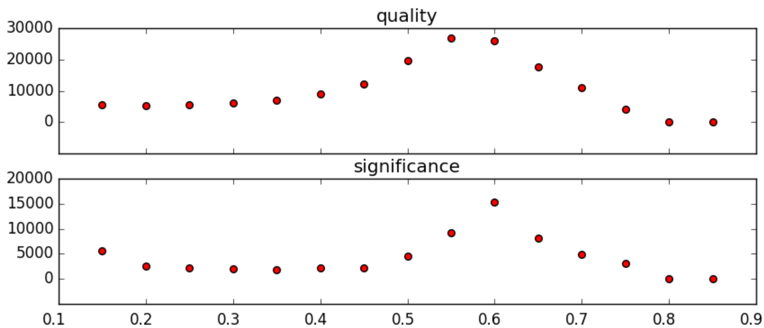


Figure 16: Quality/Significance for year 2011

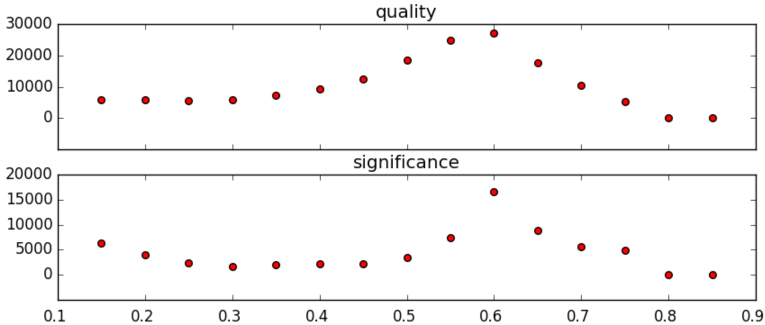


Figure 17: Quality/Significance for year 2012

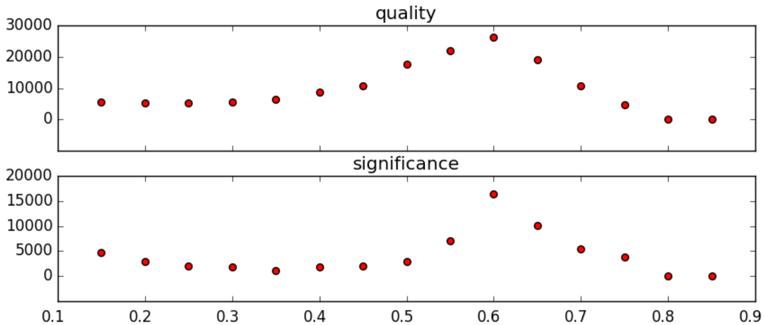
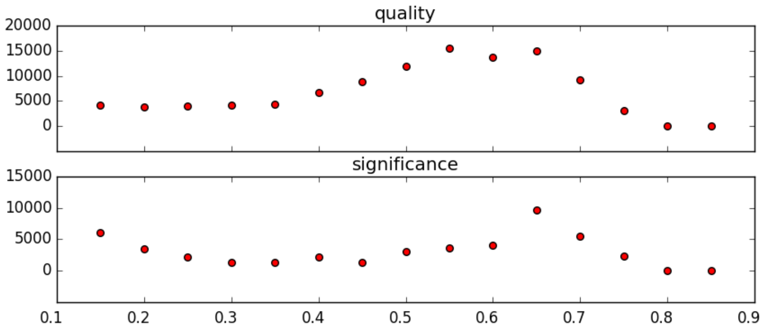


Figure 18: Quality/Significance for year 2013



B.2 Variables used in the Analysis

Table 14: Variables used in the Analysis. First column stands for the Bloomberg codes, while second column indicates the respective definitions. Items are splitted according to Balance Sheet variables and Income Statement variables.

Bloomberg Codes	**Balance Sheet Items**
BS_ACCT_NOTE_RCV	*Accounts and Notes Receivable*
BS_ACCT_PAYABLE	*Accounts Payable*
BS_ACCUM_DEPR	*Accumulated Depreciation*
BS_ASSETS_UNDER_MGMT	*Assets under Management*
BS_CASH_NEAR_CASH_ITEM	*Cash and Equivalents*
BS_COMM_LOAN	*Common Loans*
BS_CONS_LOAN	*Consumer Loans*
BS_CORE_DPST	*Core Deposits*
BS_CUST_ACCPT_LIAB_CUSTDY_SEC	*Customers Acceptance and Liabilities/Custody Securities*
BS_CUSTOMER_DEPOSITS	*Customer Deposits*
BS_DEF_TAX_LIAB	*Deferred Tax Liabilities*
BS_DEMAND_DPST	*Demand Deposits*
BS_DISCLOSED_INTANGIBLES	*Disclosed Intangibles*
BS_GROSS_FIX_ASSET	*Fixed Assets*
BS_INVENTORIES	*Inventories*
BS_LARGE_DPST	*Large Deposits*
BS_LT_BORROW	*Long-Term Borrowings*
BS_LT_INVEST	*Long-Term Investments*
BS_MKT_SEC_OTHER_ST_INVEST	*Market Securities*
BS_NON_PERFORM_ASSET	*Non-Performing Assets*
BS_OFF_BAL_COMMIT_AND_CONT	*Off-Balance Sheet Commitments*
BS_OTHER_ASSETS_DEF_CHRG_OTHER	*Other Assets*
BS_OTHER_CUR_ASSET	*Other Current Assets*
BS_OTHER_DPST	*Other Deposits*
BS_OTHER_LOAN	*Other Loans*
BS_OTHER_LT_LIABILITIES	*Other Long-Term Liabilities*
BS_OTHER_ST_LIAB	*Other Short-Term Liabilities*
BS_PFD_EQY	*Preferred Equities*
BS_RE_LOAN	*Real Estate Loans*
BS_RETAIN_EARN	*Retain Earnings*
BS_RSRV_LOAN_LOSS	*Loan Loss Provisions*
BS_SH_CAP_AND_APIC	*Share Capital & APIC*
BS_ST_BORROW	*Short-Term Borrowings*
BS_SVNG_DPST	*Saving Deposits*
BS_TIME_DPST	*Time Deposits*
BS_TOT_LOAN	*Total Assets*
BS_TOT_LOAN	*Total Loans*
INTERBANKING_ASSETS	*Interbank Assets*
Bloomberg Codes	**Income Statements Items**
INTEREST_INCOME	*Interest Income*
IS_ABNORMAL_ITEM	*Abnormal Loss (Gain)*
IS_ACT_LOAN_LOSS_NET	*Actual Loan Loss (Net)*
IS_COMM_AND_FEE_EARN_INC_REO	*IncomefromREO*
IS_INC_TAX_EXP	*IncomeTaxExpense*
IS_INT_EXPENSE	*Interest Expense*
IS_NET_NON_OPER_LOSS	*NetNon-OperatingLoss(Gain)*
IS_OPERATING_EXPN	*Operating Expense*
IS_OTHER_OPER_INC_LOSSES	*Other Operating Income Loss (Gain)*
IS_PERSONNEL_EXP	*Personnel Expense*
IS_PROV_FOR_LOAN_LOSS	*Provision for Loan Loss*
IS_TAX_EFF_ON_ABNORMAL_ITEM	*Tax Effect on Abnormal Loss (Gain)*
IS_TOT_CASH_PFD_DVD	*Total Cash Preferred Dividends*
IS_TRADING_ACCT_PROF	*Trading Account Profit (Loss)*
MIN_NONCONTROL_INTEREST_CREDITS	*Minority Interests*
NET_INT_INC	*Net Interest Income*
NON_INT_EXP	*Net Interest Expenses*
REINVEST_EARN	*Reinvested Earnings*
OTHER_ADJUSTMENTS	*Other Adjustments*

B.3 Summary statistics on geographical coverage and communities distribution

Table 15: Country Distribution. In the table we show the percentage of institutions belonging to a certain geographical area for each year. Last column gives the percentage over the entire sample. The average number of institutions is about 885. Country coverage is very stable in the dataset.

Country	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	**Average**
Africa	3%	2%	2%	2%	2%	2%	3%	3%	2%	3%	3%	3%	2%	**2%**
Arab Countries	9%	9%	8%	9%	8%	8%	8%	8%	8%	8%	9%	9%	10%	**9%**
Asia	13%	15%	15%	17%	17%	17%	17%	17%	17%	16%	17%	17%	18%	**16%**
Europe	20%	20%	20%	19%	20%	19%	19%	19%	19%	19%	19%	20%	17%	**19%**
Japan	10%	9%	9%	9%	9%	9%	9%	9%	9%	9%	9%	9%	7%	**9%**
North America	41%	41%	41%	41%	41%	41%	41%	41%	41%	41%	40%	39%	41%	**41%**
South America	4%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	**3%**
Oceania	0.4%	0.4%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.4%	**0.3%**

Table 16: Communities Distibution. First column stands for communities with abbreviations *C0*, *C1*, *C2* and *C3* meaning: *Mixed*, *US*, *JP* and *US+EU*, respectively; *Others* refers to all the other banks that are allocated to small communities (on average about 30 for each year). Column *Average* indicates the average proportions over the interval 2001-13. We can note an almost stable composition over time.

Community	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	**Average**
c0	47%	47%	47%	48%	48%	47%	48%	48%	48%	48%	48%	49%	49%	**48%**
c1	36%	36%	36%	36%	36%	36%	36%	36%	36%	36%	35%	34%	36%	**36%**
c2	11%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	8%	**10%**
c3	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	4%	**3%**
Others	4%	4%	4%	3%	3%	4%	4%	4%	4%	4%	4%	4%	4%	**4%**

Chapter 4

Peer-Group Detection of Banks and Resilience to Bankruptcy

4.1 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 one of the most important component 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 specializations. We claim that true business model identification plays an important role in shaping the risk profile of financial institutions, and therefore constitutes a precious ingredient for risk assessment.

The recent 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 financial crisis, our assessment exploits one of the largest sample of financial institutions ever tested in banking literature, which includes more than 11,000 institutions, both listed and non listed as emphasized by Köhler (2015), representing more than 180 countries. This is the result of an intensive merge and match of data from many different sources aimed to create the most comprehensive state-of-the-art global banking set. This includes a detailed list of balance sheet items collected from Bankscope and FDIC, with macro-economic and sectoral indices retrieved from BIS, Datastream, OECD and WorldBank, that will be used to characterize and control banks' business models. For the distress even list, we merge 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 level, 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[1] of peer group assessment, which defines groups of institutions adopting the same business model. We borrow a classification method well-established in the complex sys-

[1]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.

tem field where *i)* similarities among pairs of financial institutions are measured by the cosine similarity of their vectors of attributes (Dongen and Enright 2012), representing a detailed list of balance sheet items, 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 are consistent with the three classic business model categories discussed in previous works (e.g. Ayadi et al. 2012; Demirgüc-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 categories is mainly based on their funding characteristics. However, due to the mix features of the business models we detected, we label them according to the distinctive balance sheet characteristic that best differentiate each model from the others. Among the specific business models that we find, three of those 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 banks holdings. The second is another wholesale-oriented model represented by a solid composition of US commercial banks and European institutions 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. Furthermore, 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 groups emerging for few years only and with fewer institutions: a *Long Term* business model characterized by dominant long term funding, commercial investments and composed by mainly European institutions (in particular 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 nondeposit 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 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 indicator (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 does not suffer from different accounting standards.

To assess the performance of our classification model against the other state-of-the-art methods we exploit a variety of tests. First, we perform a battery of non-parametric tests and multiple pairwise post-hoc comparisons on peer groups differences in their business models features to confirm that each banking group is statistically different to the others. Second, we compare the results of our approach with other methods adopted in literature, such as the wide-diffused Ward clustering algorithm (Ward and Joe 1963), and the direct classification provided by Bankscope and used in Köhler (2015). We show that our classification provides a better clustering identification, thus supporting the use of the resulting clusters as peer groups/business models. Last, we focus on the stability of peer groups' membership over time by analysing 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. Our switching model analysis suggests that both size of the institutions and macro-economic environments might play a role in switching business model, although estimates clearly highlight that those institutions that are more peripheral/dissimilar within their peer group tend to be more prone to change business model.

The last part of the study investigates the contribution of specific business models to the assessment of the risk of distress. We built a comprehensive list of global distress 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 2012). We regress these events against financial statement ratios (i.e. proxies for CAMELS) and controlling for macro and sectoral effects using a rare-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; Demirgüc-Kunt and Huizinga 2010; Köhler 2015; Mergaerts and Vander Vennet 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 discover different driving forces 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 those business models with a much lower proportion of stable funding than the deposit-oriented group. Second, ROA impacts positively deposit-oriented models and negatively the wholesale ones, suggesting that limited asset diversification of the deposit oriented institutions would force them to concentrate their investments on fewer product in the pursuit of higher returns, with overall higher risk compared to the wholesale. Opposite result for ROE suggesting that leverage may exacerbate the deposit oriented institutions stable funding and therefore improve resilience while wholesale oriented would pay the price of more unstable non-deposit funding growth in the pursuit of better returns on equity. 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. We confirm that the higher the volatility of business model adopted, the higher the likelihood of distress. 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.

Drawing on our results, we elaborate policy recommendations. We find different risk drivers for banks adopting competing business models, 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 4.2 discusses the main literature on banking business models[2] classification and contributions to performance and resilience. The dataset constructed for the empirical analysis is described in Section 4.3. Section 4.4 introduces the methodology for banking peer groups classification, whereas the discussion of their main features and geographical compositions is presented in Section 4.5. The inter-temporal analysis of the behaviour of business models is analysed in Section 4.6. Section 4.7 provides a detailed risk assessment of the resilience of institutions to the distress events of 2008-10. Final conclusion and remarks are given in Section 4.8.

[2]In the paper we se interchangeably *business model* and *peer group* as the result of the clustering procedure. Due to the characteristics of the dataset, we also use *banks* and *institutions* as synonimous.

4.2 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, Demirgüc-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. 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 and Laeven and Valencia 2012), 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 from Bankscope up to 2004, which is discussed in Section 4.3, 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 resiliance 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 therefore 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[3] 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. 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 dataset 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 iden-

[3] A detailed review of banking failure prediction model is given in Demyanyk and Hasan (2010).

tification. 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-based models. Köhler (2015) confirm results shown in Demirgüc-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 4.5). 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 veru stable over time). We also perform an intertemporal analysis by employing a switching 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 the largest sample ever tested in literature. 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 datasets. Comparisons between our approach and the above methods are discussed in Section 4.5.

4.3 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[4]. In addition to that, the coverage of institutions in Bankscope improves substantially in 2005 with the inclusion of many European banks[5]. To investigate the presence of similarities across institutions in different countries we consider a global sample covering more than 180 countries (more than 100 countries with at least 10 institutions). Finally, our clustering 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 present in Duprey and Lé (2014).

4.3.1 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 17 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[6]. Annual mean values suggest that on aver-

[4]For a timeline of the relevant changes in IAS-IFRS, the interested reader can refer to FASB (2016).

[5]For some countries the number of institutions that are present increases significantly from 2004 to 2005. This is the case for instance 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.

[6]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.

age 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 17) 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 share of the respective total assets[7].

4.3.2 Distress Events

In Table 18 we show 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).

[7] 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.

Table 17: 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.

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

Table 18: 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.
Bankruptcy	*This event occurs if the net worth of the bank falls below a country specific regulatory threshold*	*Bankscope*	**8**
In Liquidation	*This event is related to the sale of the bank by the liquidator as per the guidelines of the country regulations*	*Bankscope*	**17**
Default	*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*		**26**
	•*Short-Term rating NP (Not Prime) and Long-Term rating Caa (speculative or poor standing) or below*	*Moody's*	
	•*Short-Term rating SD (Selective Default) or D (Default) and Long-Term rating CCC (vulnerable) or below*	*Standard & Poor's*	
Distressed Merger	*Merger (from Bankscope) which involves a bank with distressed conditions*		
	Coverage Ratio in $t-1<0$		**19**
	•*(Equity + Reserves for NPLs - Total Impaired Loans)/ Total Assets* <0	*Bankscope*	
	Short or long ratings below vulnerable state in t		**3**
	•*Short-Term rating NP (Not Prime) or Long-Term rating Caa (speculative or poor standing) or below*	*Bankscope/Moody's*	
	•*Short-Term rating SD (Selective Default) or D (Default) or Long-Term rating CCC (vulnerable) or below*	*Bankscope/S&P*	
Public Bailout	*State Aid from Governments, such as: Nazionalization, Recapitalization and Assets Guarantees and Purchases*	*Laeven and Valencia (IMF, 2013)*	**77**
Other	*US Failed bank list from the Federal Deposit Insurance Corporation*	*FDIC*	**63**

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. For this reason, we introduce additional distress definitions. 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. In addition, we consider as distressed mergers those cases where the institutions present a rating indicating a vulnerable state. Finally, we enrich the dataset of distressed institutions by including the information of public bailouts (Laeven and Valencia 2012) and, for the US perimeter, we integrate Bankscope with bankruptcy information from the Federal Deposit Insurance Corporation.

4.3.3 Indicators of Vulnerability

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 are 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 wider coverage across banks balance sheets. 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 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 financials obligations and thus are perceived as less risky. Finally, the sensitivity to market risk is measured by the share of securities to total assets (Total security 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.

Table 19: 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	*Equity/Total Assets*	*Bankscope*
Capital Funding Ratio	*(Common Equity + Subordinated Borrowing)/Total Assets*	*Bankscope*
Roa	*Return on Average Assets*	*Bankscope*
Roe	*Return on Average Equity*	*Bankscope*
Cost to Income Ratio	*Operating Expenses/Operating Income*	*Bankscope*
Net Interest Margin	*Net Interest Revenues/Total Earning Assets*	*Bankscope*
Interest Expenses to Total Liabilities	*Total Interest Expenses/ Total Liabilities*	*Bankscope*
Liquid Assets to Short - Term Funding	*(Cash and other liquid assets)/(Deposits and ST Funding)*	*Bankscope*
Deposits to Total Funding	*(Deposits and ST Funding)/Total Funding*	*Bankscope*
Total Securities to Total Assets	*Total Securities/Total Assets*	*Bankscope*
GDP per capita	*Annual percentage growth rate of GDP per capita*	*World Bank*
Inflation	*Consumer prices (annual %)*	*World Bank*
House Price	*Real house prices*	*OECD/BIS*
Unemployment	*Share of the labor force that is without work but available for and seeking employment*	*World Bank*
FDI-Inflows	*Net inflows (% of GDP)*	*World Bank*
FDI-Outflows	*Net outflows (% of GDP)*	*World Bank*
Central Gvt. Debt	*Entire stock of direct government fixed-term contractual obligations (% of GDP)*	*World Bank*
Gvt. Long-Term Yield	*Long-term government bond yield*	*Datastream*
Bank NPLs to Gross Loans	*Nonperforming loans divided by the total value of the loan portfolio*	*World Bank*
Credit to Non-Financial Sector	*Banks domestic credit to non financial sector*	*BIS*
Market Index	*S&P Global Equity Indices (annual % change) or, alternatively, FTSE Indices*	*World Bank/Datastream*
Sector Index	*FTSE Financial Indices*	*Datastream*
Stock Traded	*Number of traded shares multiplied by their respective matching prices*	*World Bank*

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 dataset 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, bank non-performing loans to gross loans, central government stock of debt (% of GDP), long-term government bond yield, financial sector market returns, stock market returns and the amount of traded stocks. Moreover, we take into account macroeconomic conditions by controlling for GDP per capita growth, inflation, house price, unemployment rate and foreign direct investments (in and out). 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.

4.4 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 which, although in line with the Ward algorithm, is designed to address in a more appropriate and elegant way the multi-dimensionality issues, particularly severe in a large and sparse sample like ours (see Table 17). 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.

4.4.1 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[8] is computed as follows:

$$CosSim(x,y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} = \frac{< x, y >}{||x||||y||} \tag{4.1}$$

Once we have measured the pair-wise similarities among institutions we apply a hierarchical clustering algorithm which identifies groups by

[8]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.

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:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4.2}$$

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 which has received an increasing interest in complex systems literature[9]. 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. ΔQ) by moving an isolated institution i into a cluster c can be measured as follows:

[9]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).

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (4.3)$$

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 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 pairwise similarity between (*i,j*) as $1 - \theta_{i,j}$, so that pairs of institutions which are very similar receive higher weights.

There are a couple of technical issues that are worth mentioning. 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[10]. We recall that edges with higher values stand for higher

[10]In particular, we filter the system according to thresholds from 0.7 to 0.5 using a de-

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.

Appendix C.2 shows results from the non-parametric equality of medians tests (Kruskal-Wallis) that it 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 (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 which 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 20 compares the *Louvain* results with those obtained 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

creasing step equals to 0.025 and, for each year, we select the threshold which maximizes the significance of the configuration.

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 dataset. 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 17 and we fill missing values (NAs) using four criteria to guarantee that the complete dataset is used. 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) 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 definition the cosine similarity treats them as unknown and base the similarity analysis on available data only.

For the *Louvain* method and the direct classification we use the resulting clusters from the respective approaches and, to enhance comparability, we compute clustering validation measures filling missing values using the same four criteria as seen above.

4.4.2 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:

$$Pr(Y_i = 1|\mathbf{x}_i) = \Lambda(\mathbf{x}_i\beta) \tag{4.4}$$

where Λ 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 4.3.2). 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 4.3.3). 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.

4.5 Global Business Models

This Section discusses the outcomes of our classification approach by investigating the business models we find, their characteristics and geographic compositions as well as their evolution over time. The first subsection provides a comparison between our clustering approach and both the Ward method and the direct classification. We then show 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.

4.5.1 Clustering Validation

First of all, we run some tests 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 4.4.1), 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 dataset 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[11]. 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. 4.1). 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

[11]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.

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 20 provides performance measures of each of the three clustering methodologies under four main assumptions for filling missing values (see Section 4.4.1 for details): [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.

Table 20: 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.

	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

Source: Authors' own calculation

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[12] may adopt very different business models. Therefore, this classification may not provide a useful indication of the activities they 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 C.3, where we provide evidences on differences in balance sheet dimensions across business models by means of non-parametrically tests.

[12]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.

4.5.2 Business Models Characteristics

We focus here on the outcomes of our classification approach[13]. Tables from 21 to 23 present summary statistics for the main economic aggregates used to characterise and distinguish business models under different time intervals[14]. 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 31 (and discussed in detail in Section 4.6), confirms our findings of stable peer groups and supports 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[15] 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 asset side (Ayadi et al. 2012; Roengpitya et al. 2014). Detailed individual structures of each business model and comparisons with direct classifications are discussed in the next following subsections (we continue to present direct classification together with our clustering results to provide a helpful parallel to common banks de-

[13]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 subsection 4.5.1), and in particular the parallel between our approach and Ward, to future studies.

[14]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 C.2).

[15]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 asset side (Retail Loans), we prefer to use this terminology as customer deposits are what characterized the retail-funded institutions.

scriptions). Three main business models are consistent throughout the entire reference period 2005-14 and accounting for the largest number of institutions. The first of these models is characterised by the largest wholesale funding and well diversified loans investments with decent exposure to interbank activities. Many Russian commercial banks adopt this business model along with US bank holdings. We name this model *Wholesale*[16]. 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. Due to the nature of their loan investments, we name this model *Commercial*. The last model is based on traditional customer deposits funding (the largest across models) invested in loans, mainly commercial. Many US and Japanese commercial banks adopted this model. We name this model as *Saving* group. 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[17].

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 19 shows the peer groups at the beginning of our sample period. On top of our three core models introduced above, our classification 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

[16]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 direct method.

[17]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.

Table 22. 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. Due to the close similarity to the *Commercial* model with regards to their asset 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 pictured in Figure 20. 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 saving banks and US local bank holdings.

Figure 19: Peer Groups in 2005-2006. Visualization of the peer groups adopted in the period 2005-2006. 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. The geographic representation of the main banking specializations are reported in the squared boxes, where the specialization acronyms from Bankscope stand for Bank Holdings (BH), Commercial (C), Cooperative (Coop), Investment (I), Private Bank & Asset Management (PB&AM), Real Estate & Mortgage (RE&M), Saving (S).

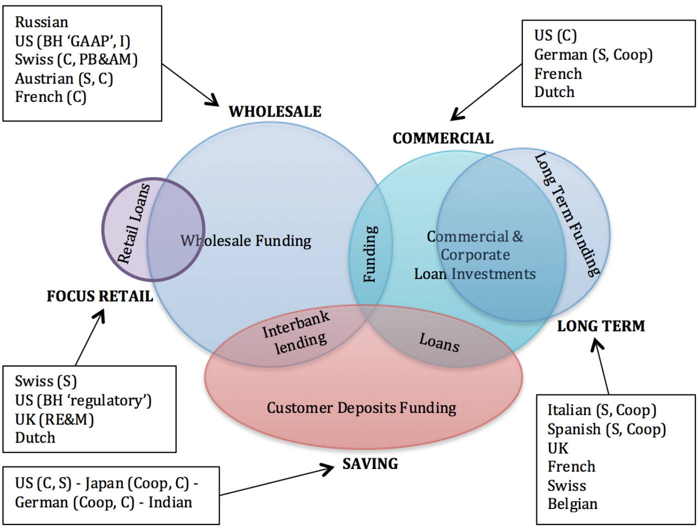
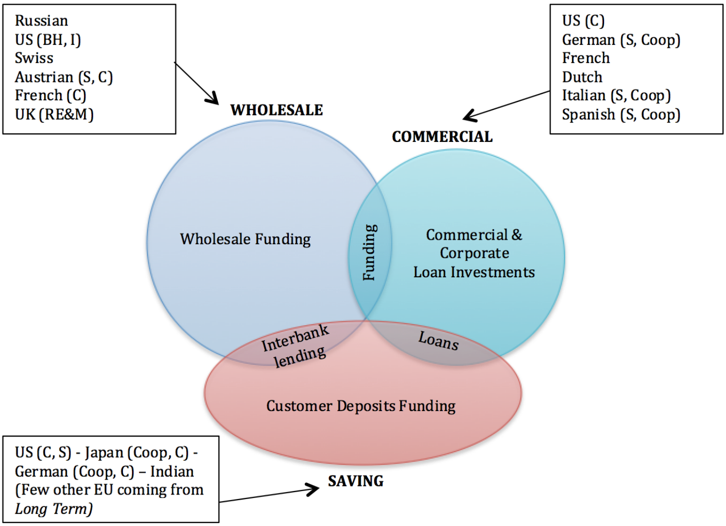
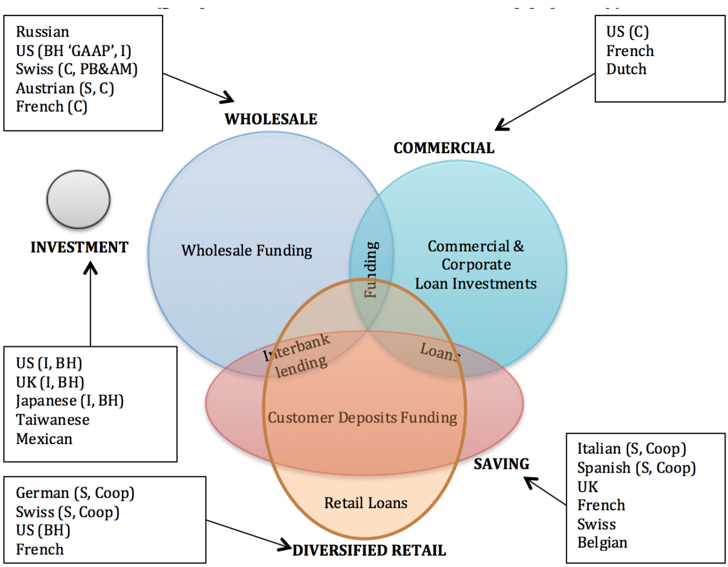


Figure 20: Peer Groups in 2007. Visualization of the peer groups adopted in 2007. Intersection sets represents the common characteristics of the business models while the distinctive business model characteristics are listed in the centre of each group set. Sizes of the groups represents the popularity of business models measured by the total number of banks. The geographic representation of the main banking specializations are reported in the squared boxes, where the specialization acronyms from Bankscope stand for Bank Holdings (BH), Commercial (C), Cooperative (Coop), Investment (I), Private Bank & Asset Management (PB&AM), Real Estate & Mortgage (RE&M), Saving (S).



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. Figure 21 depicts this group composed by mainly German and Swiss saving and cooperative institutions which clusterised together after 2008 in this new business model that persisted thereafter. Statistical information and comparisons with the three core models are given in Table 23. Similar to the 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).

Figure 21: Peer Groups in 2008-2014. Visualization of the peer groups adopted in the period 2008-2014. Note that *Investment* group is adopted in 2012 and 2014 only. 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. The geographic representation of the main banking specializations are reported in the squared boxes, where the specialization acronyms from Bankscope stand for Bank Holdings (BH), Commercial (C), Cooperative (Coop), Investment (I), Private Bank & Asset Management (PB&AM), Real Estate & Mortgage (RE&M), Saving (S).



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 C.3.4), 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. We named this model *Investment*, which appears only in 2012 and 2014 (Figure 22 shows peer groups' features before and after the onset of the crisis).

Table 21: Peer Groups Economic Features. We report average 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 C.2. Estimates are computed over the interval 2005-14. Last row provides summary statistics for Total Assets (in USD Billion).

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

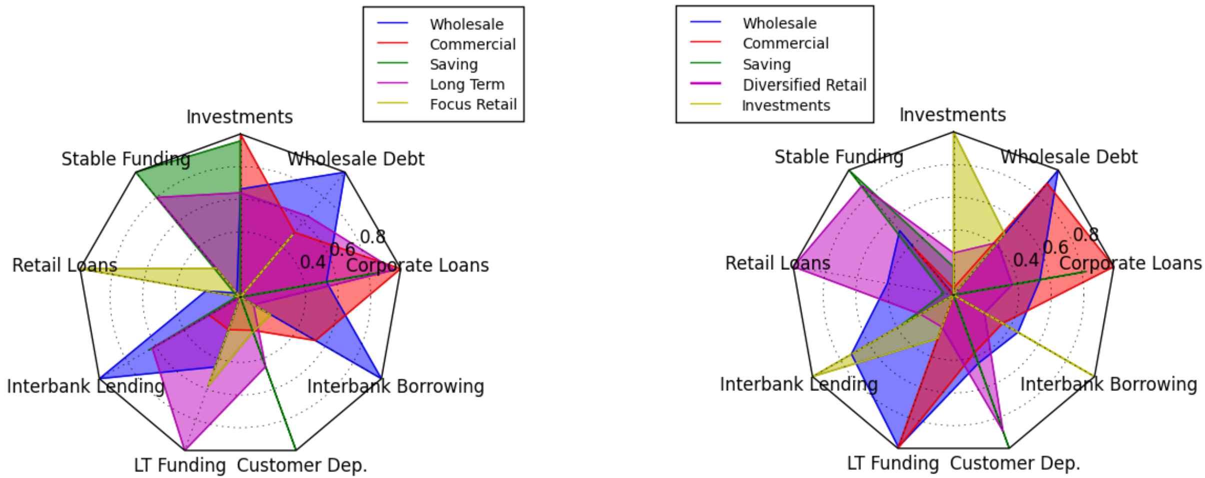
Table 22: Peer Groups Economic Features - Pre Crisis. We report average values for aggregated balance sheet variables standardized by total assets for groups Wholesale, Commercial, Saving, Long Term and Focus Retail. Column *Average (I-III)* refers to the average values within the sample composed by 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 C.2. 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).

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

Table 23: Peer Groups Economic Features - Post Crisis. We report average 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 C.2. 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).

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

Figure 22: Peer Groups Main Features. Radar plot on the left shows the main balance sheet features for each peer group in the period 2005-07, while plot on the right exhibits the main characteristics resulting in the interval 2008-14. Values are normalized within each measure and period to better visualize the main features of the peer groups and their relative differences. For variables' definitions see Appendix C.2.



Wholesale

The *Wholesale* business model is the most popular model in our sample, accounting on average 3,490 institutions in the period 2005-14. This model is characterised by the largest proportion of wholesale funding, 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 asset 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%). The above figures are consistent with the "Wholesale-funded"[18] model discovered in Roengpitya et al. (2014) on a much smaller global sample. We also note substantial similarities with the "Wholesale"[19] 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 activity, the *Wholesale* model presented here reflects interesting non-traditional banking activities that could jeopardize stability in distressed scenarios (see e.g. Lozano-Vivas and Pasiouras 2010).

This model reaches the peak of popularity in 2007 with 4,108 institutions, as displayed in Table 24. 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 4.5.2). Swiss insti-

[18]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 21. Their interbank composition is less prominent than ours, but still consistent across other models as we will show in the next subsections.

[19]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.

tutions 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 4.5.2. 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 4.5.2. 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 23), 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 which represent almost one third of the total peer group's assets. 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). Finally, we anticipate that popularity and low percentage of stable funding relative to total assets for this business model come at a price. Almost half of the distress events we collect globally come from institutions adopting this model at some points in the period 2005-07 (precisely: 78 in 2005, 75 in 2008 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 provide clear evidence that being bigger relative to average peer group members, even from the same country, exacerbates vulnerability and eventually distress (see Section 4.7).

Table 24: 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.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
Obs.	3221	3390	4108	3270	3116	3484	3608	3626	3424	3653	**3490**
Country											
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%**
Specialization											
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%	5%	6%	5%	2%	4%	**6%**
Other	19%	19%	21%	20%	18%	19%	20%	19%	21%	21%	**20%**

Source: Bankscope, authors' own calculation

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.

Figure 23: Total Assets Distribution: *Wholesale* Model. Red curve represents the time series of average total assets (in US Billion) for all Wholesale 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 calculation

Commercial

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 21 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 25 confirms the dominant position of US institutions[20], 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 4.5.2. 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 4.5.2) 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 Ahorros) also migrated into this model in the years 2008-09.

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, US institutions adopting wholesale-oriented business models are bank holdings and national commercial banks, that were not part of the sample in Roengpitya et al. (2014). The extension of the dataset and the use of our indirect approach can, therefore, clarify banking activities across a wider range of institutional specializations and depict a bigger picture to allow for a more

[20]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.

accurate analysis of business models both within and across countries.

Table 25: 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.

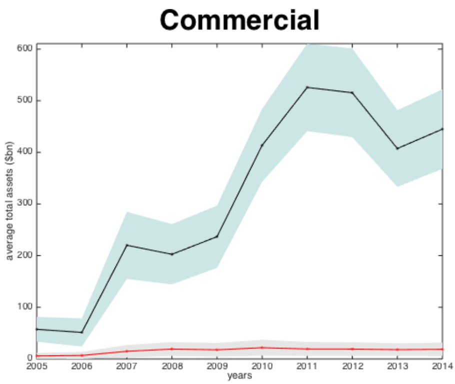
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
Obs	2592	2568	3861	3003	2877	2760	2692	2125	2135	2390	**2700**
Country											
UNITED STATES OF AMERICA	23%	23%	15%	20%	24%	23%	22%	28%	27%	23%	**23%**
GERMANY	59%	59%	38%	11%	4%	4%	5%	4%	4%	4%	**19%**
ITALY	1%	2%	16%	22%	22%	23%	23%	26%	25%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

Source: Bankscope, authors' own calculation

Looking at the distribution of total assets (see Figure 24), 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 the one 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).

On the risk profile, we observe 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 belonging always to the *Commercial* group during the interval 2005-07) than the *Wholesale*. 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.

Figure 24: Total Assets Distribution: *Commercial* Model. Red curve represents the time series of average total assets (in US Billion) for all Commercial 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 calculation

Saving

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. 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 latter, in combination with interbank borrowings, makes those institutions the only interbank net lenders among the three core models as shown in Table 21. 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 will be discussed in Section 4.5.2. In fact, the latter shows a better European representation than the *Saving* model and may 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 with a well-diversified assets side. Roengpitya et al. (2014) find that their retail-funded model, closely related to our *Saving* model[21], presents less volatile earnings compared with the other models, although it is not as cost-efficient as the wholesale-funded one.

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).

[21]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.

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.

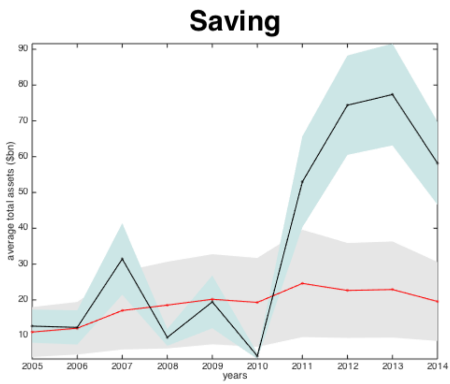
Table 26: Saving. Summary descriptive statistics over the interval 2005-14. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period.

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**Average**
Obs	1946	1952	2070	1708	2125	2009	1778	1901	1849	1807	**1915**
Country											
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%**
Specialization											
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%**

Source: Bankscope, authors' own calculation

With more than \$20tn worth of total assets (see Figure 25), *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. The *Saving* group differentiates itself from the other two core groups (and from any other wholesale-oriented as we will see in next subsections) also on the risk attitude of the institutions. In fact, the is no evidence of higher vulnerability of those on the right side of the distribution of assets relative to their peers. Average sizes of distressed institutions during the period 2008-10 in this group are 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 issue will be discussed in detail in Section 4.7.

Figure 25: Total Assets Distribution: *Saving* Model. Red curve represents the time series of average total assets (in US Billion) for all Saving 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 calculation

Diversified Retail

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, 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. These results suggest a substantial restructuring process implemented by German and Swiss institutions at the peak of the financial crisis, aiming at improving their stable funding, which indeed increased from about 53% (average of *Wholesale* and *Commercial* models) to 77%. The result is an hybrid liability structure that incorporates characteristics of both *Saving* and *Wholesale* 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%). Following Köhler (2015), Section 4.7 will discuss and compare the impact of variations of noninterest income and non-deposit funding to the risk of distress by focusing on the changes of business models prior to the crisis between deposit- and wholesale-oriented ones.

Regarding the geographic composition of this group, Table 27 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 4.5.2) group (60%).

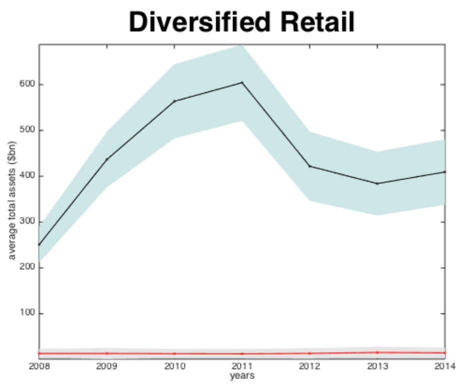
Table 27: Diversified Retail. Summary descriptive statistics over the interval 2008-2014. Column *Country* reports the main countries of membership; *RoW* refers to the Rest of the World. Column *Specialization* stands for the classification provided by Bankscope; *Others* summarizes all the other marginal specializations. Column *Average* refers to the average values over the reference period.

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

Source: Bankscope, authors' own calculation

The popularity of this new peer group is also confirmed by the share of the total assets (see Figure 26) 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).

Figure 26: 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 calculation

Long Term

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.

Table 28: 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.

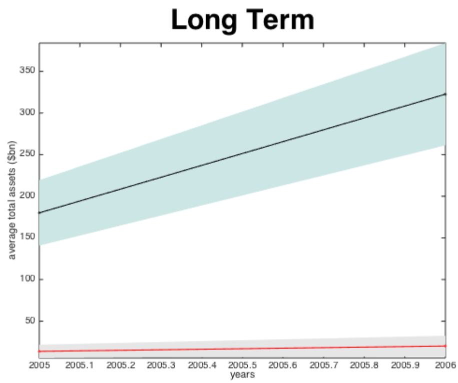
	2005	**2006**	**Average**
Obs	1416	1534	**1475**
Country			
ITALY	43%	40%	**41%**
SPAIN	7%	6%	**7%**
SWEDEN	5%	5%	**5%**
GERMANY	2%	4%	**3%**
UNITED KINGDOM	3%	3%	**3%**
FRANCE	2%	3%	**2%**
SWITZERLAND	0%	2%	**1%**
RoW	37%	37%	**37%**
Specialization			
Cooperative Banks	38%	36%	**37%**
Commercial Banks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

Source: Bankscope, authors' own calculation

Looking at Table 22 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 27). 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). The risk profile here 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 sizes.

Figure 27: Total Assets Distribution: *Long Term* Model. Red curve represents the time series of average total assets (in US Billion) for all Long Term 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 calculation

Focus Retail

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. 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[22].

[22]For a deeper analysis on banks capital requirements, see reports in *http://www.bis.org/publ/bcbsca.htm*.

Table 29: 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.

	2005	2006	Average
Obs	606	463	**535**
Country			
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%**
Specialization			
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%**

Source: Bankscope, authors' own calculation

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 4.5.2 where they finally 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.

Figure 28: Total Assets Distribution: *Focus Retail* Model. Red curve represents the time series of average total assets (in US Billion) for all Focus 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 calculation

This peer group does not exceed in assets size (see Figure 28) 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). As for the other wholesale-oriented models 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.

Investment

The last business model we present here is the *Investment* model, which only runs in 2012 and 2014. Table 23 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)[23] and Ayadi et al. (2012)[24].

[23]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.

[24]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.

Investment model is also characterized by a 20% interbank lending, placing institutions in this group among the largest net borrowers in the interbank market.

As we expected, the geographic representation of the *Investment* group presented in Table 30 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.

Table 30: 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.

	2012	2014	Average
Obs	310	226	**268**
Country			
Country			
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%**
Specialization			
Investment Banks	28%	36%	**32%**
Commercial Banks	24%	20%	**22%**
Securities Firm	23%	21%	**22%**
Bank Holding & Holding Companies	11%	9%	**10%**
Other	15%	14%	**14%**

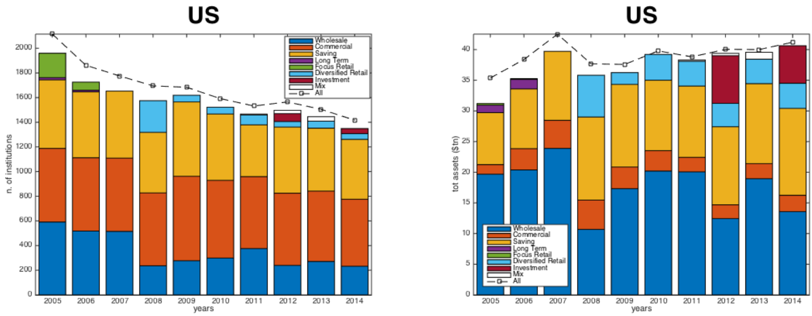
Source: Bankscope, authors' own calculation

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 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.

4.5.3 Geographic Map of Business Models

This final section extends the discussion of the business models of Section 4.5.2 by focusing on banking activities of the main countries. Figures 29-38 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 to classify despite the lack of consistency of data availability for some institutions that were excluded from our classification (see Footnote 6).

Figure 29: US Business Models. The plot shows the composition of banking business models in US 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 dataset.

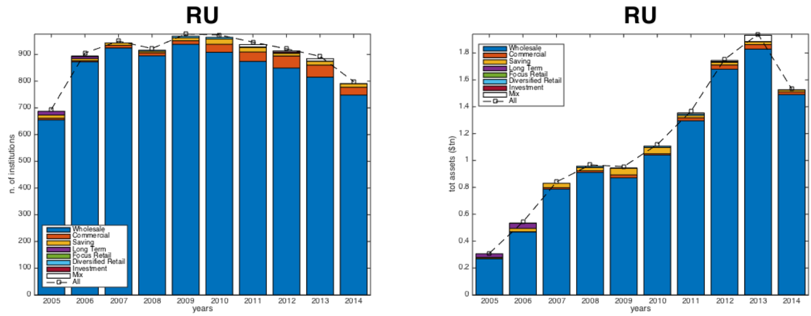


The US banking system is the largest system worldwide with total assets exceeding \$40tn in 2007[25]. US banks experienced one main falls at \$36tn right at the peak of the financial crisis in 2008 with a sluggish recovery affected by the US debt ceiling crisis in 2011 (Figure 29 (right) dashed line). At the bank level, we note a constant reduction of number of banks, due to defaults and merges caused by the financial crisis,

[25] Values are the representation of Bankscope database, which covers more than 90% of the total assets values VIP (2011).

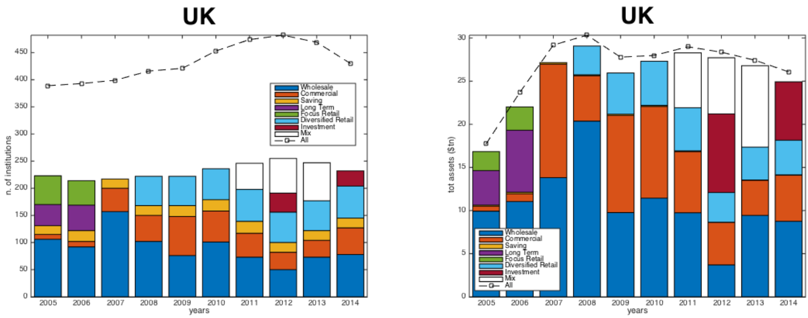
from more than 2000 institutions in 2005 to 1400 in 2014. Although the US banking system appears to be very well diversified and balanced in terms of business models, the fall in numbers has mainly impacted banks adopting the *Wholesale* model, the largest in terms of total assets (Figure 29 (right)). 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 activity 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.

Figure 30: Russian Business Models. The plot shows the composition of banking business models in Russia 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 dataset.



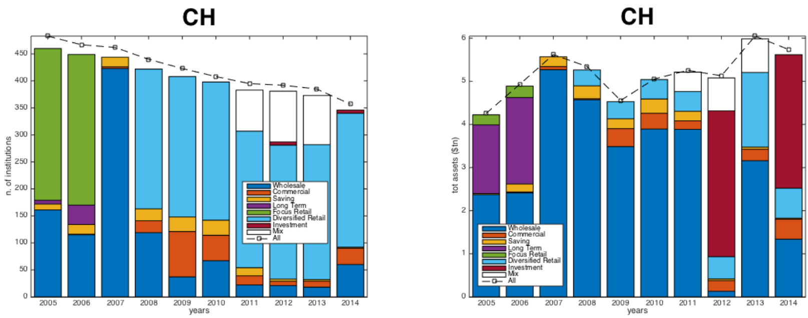
Opposite scenario for the Russian banking system, characterized by almost all *Wholesale* banks, with few *Commercial* and *Saving* institutions emerging in the post financial crisis. Although the number of banks fluctuated among 900 institutions in Figure 30 (left), we note a speedy growth in total assets from $300bn in 2005 to almost $2tn on 2013 in Fig 30 (right). 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).

Figure 31: UK Business Models. The plot shows the composition of banking business models in UK 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 dataset.



In Europe, the UK banking system is the second largest in terms of total assets with values reaching almost $30tn in 2008 in Figure 31 (on the right), even though those assets are owned by a very small number of institutions (380 in 2005 raising up to almost 500 in 2012 as shown in Figure 31 left). We note that a decent share of very small UK institutions were excluded by our classification approach due to lack of data coverage across balance sheet variables (see Footnote 6). This is confirmed by the fact that the total assets coverage is extremely high (over 90%) as for all other countries. The financial crisis hit the UK banks mainly from 2009 with a loss of 10% of assets that deepened in recent years. Like the US system, UK banks tend to be quite diversified in terms of business models, with predominant wholesale-oriented funding structure. Indeed, only few banks were adopting the *Saving* model ($< 10\%$) whereas the others were mainly on *Wholesale* and *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 *Commercial* model in 2007. Part of them 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.3.4.

Figure 32: Swiss Business Models. The plot shows the composition of banking business models in Switzerland 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 dataset.



The case of Swiss banks is interesting as they experienced a big transformation after the 2008 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 32 (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 32 (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).

Among the main Eurozone countries plotted in Figure 33, 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 4.5.2, German banks along with small Swiss cooperative and saving banks that predominately used to adopt the *Commercial* model, 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 *Wholesale* activity towards *Commercial* and, marginally, *Diversified Retail*, 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 (Bentolila et al. 2012).

In the Asian area pictured in Figure 34, 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 10 years. 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 35) compared to the East European countries that share a more European business model pattern (more diversified with a large share of the *Commercial* business model, Figure 36). Finally, we report quite mixed trends in both in the Islamic countries (Figure 37) as well as the major offshore financial centres Figure 38.

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* models.

Figure 33: 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 dataset. Countries are Germany (DE), France (FR), Spain (ES), Italy (IT).

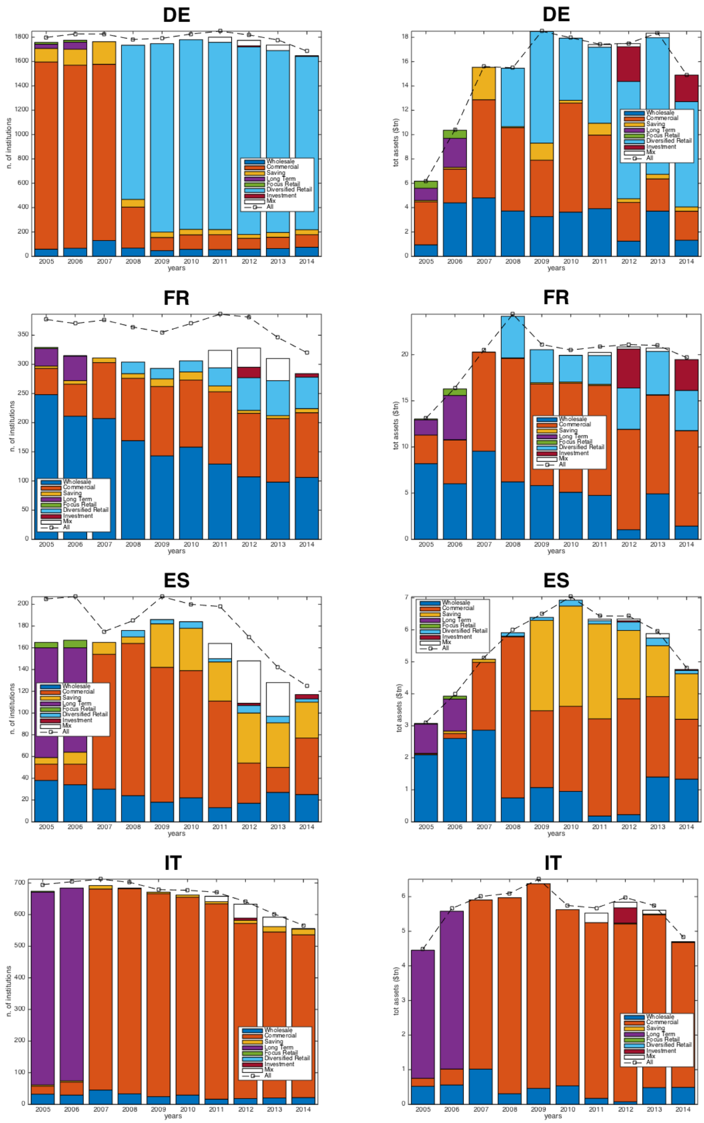


Figure 34: Major Asian Business Models. The plot shows the composition of banking business models in the Asian area in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope dataset. Countries are China (CN), Honk Kong (HK), Japan (JP), India (IN).

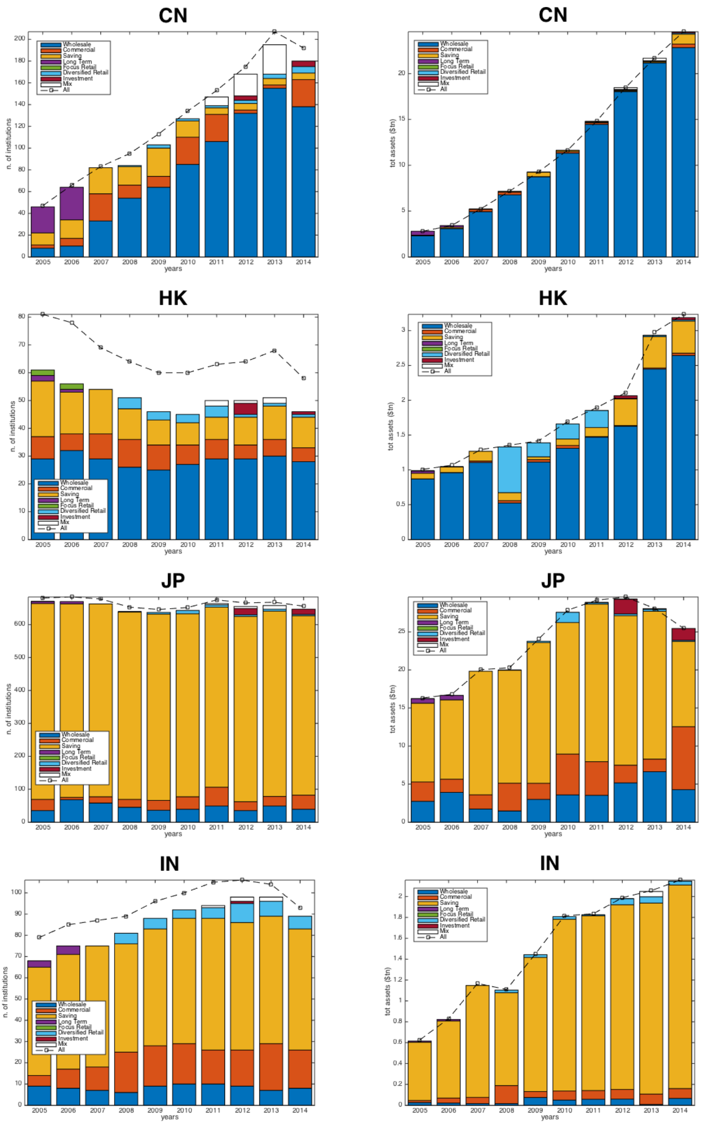


Figure 35: Major Latin America Business Models. The plot shows the composition of banking business models in the Latin America area in terms of number of banks (left) and total assets (right, in $tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope dataset. Countries are Mexico (MX), Brazil (BR), Colombia (CO), Argentina (AR).

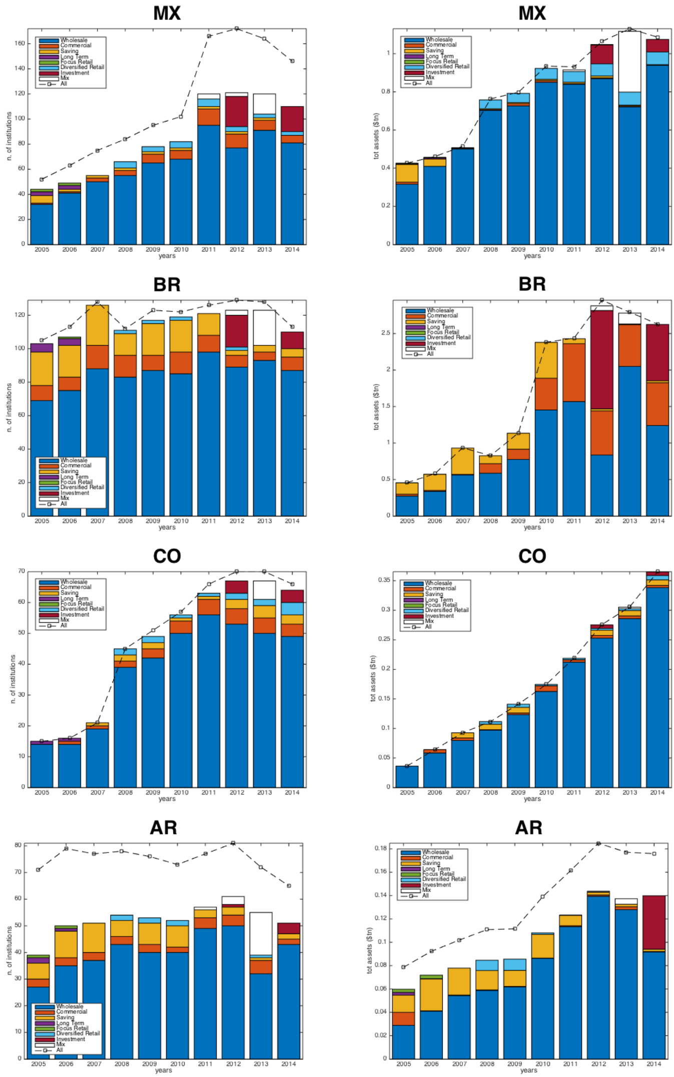


Figure 36: Major East Europe Business Models. The plot shows the composition of banking business models in the East Europe area in terms of number of banks (left) and total assets (right, in $tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope dataset. Countries are Poland (PL), Czech Republic (CZ), Hungary (HU), Slovakia (SK), Slovenia (SI).

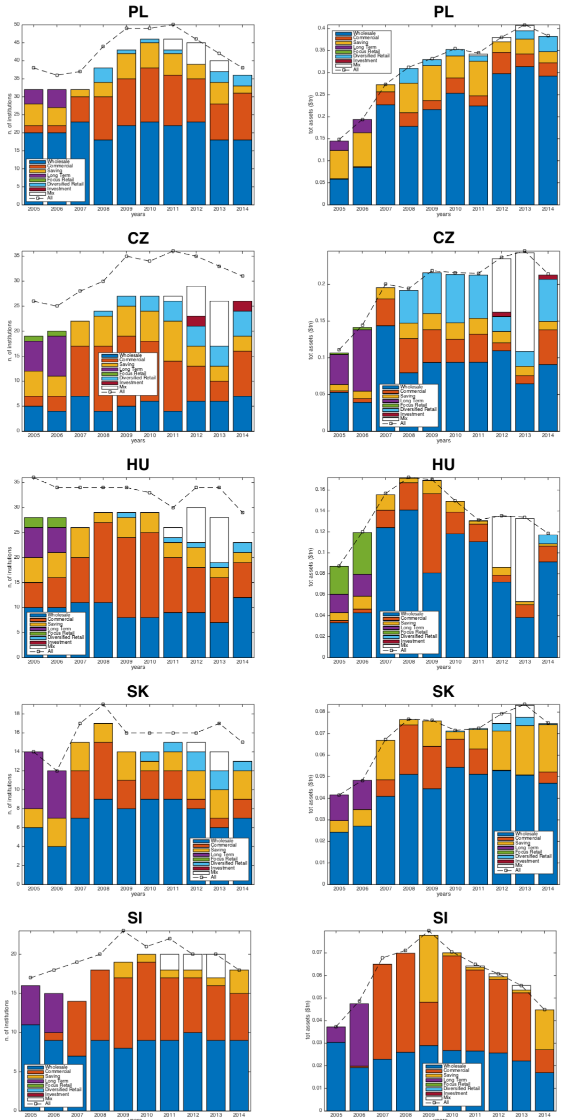


Figure 37: Major Islamic Countries Business Models. The plot shows the composition of banking business models in the Islamic area in terms of number of banks (left) and total assets (right, in $tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope dataset. Countries are Saudi Arabia (SA), Bahrain (BH), Kuwait (KW), United Arab Emirates (AE).

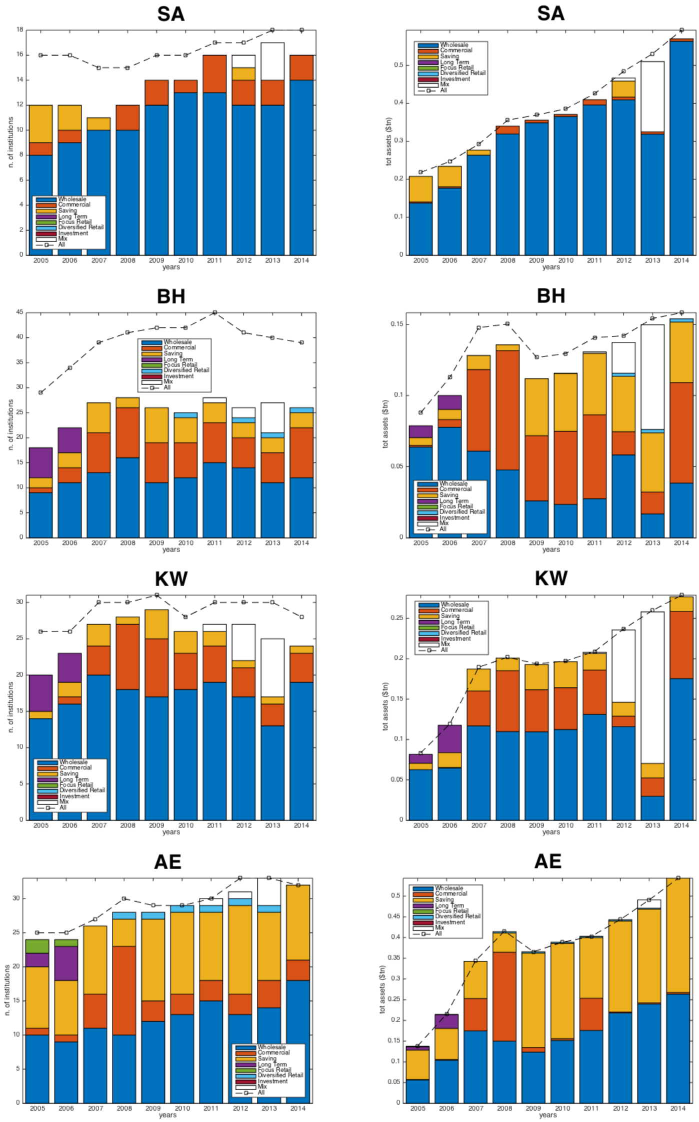
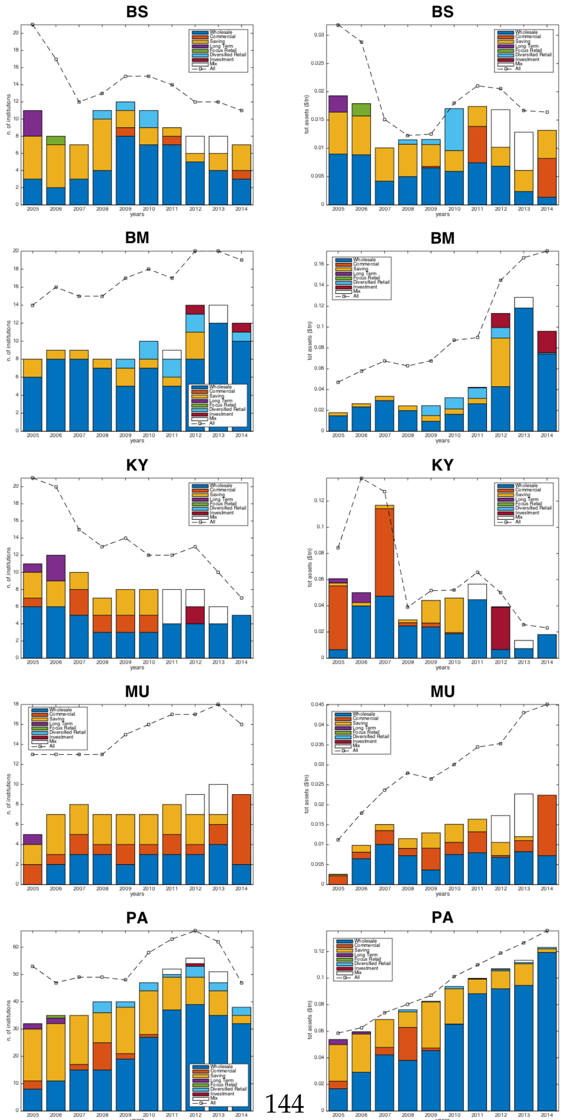


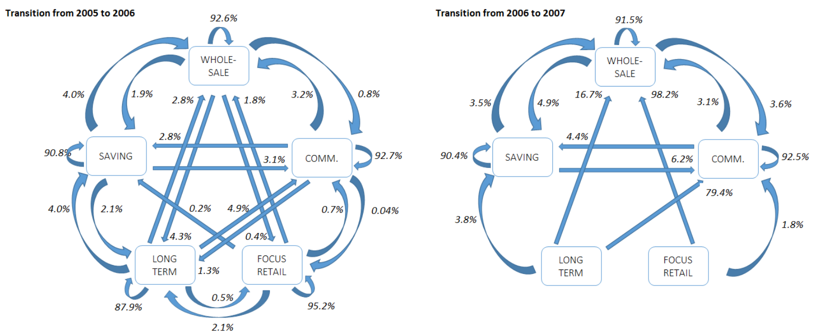
Figure 38: Major Offshore Financial Centres Business Models. The plot shows the composition of banking business models within offshore financial centres in terms of number of banks (left) and total assets (right, in \$tn) from 2005 to 2014. The dashed black line plots the number of banks (left) and aggregate total assets (right) of all Bankscope dataset. Countries are Bahamas (BS), Bermuda (BM), Cayman Islands (KY), Mauritius (MU), Panama (PA).



4.6 Switch Analysis

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. 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 C.3 for a detailed analysis of year-by-year representation of the business models).

Figure 39: 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.



As shown in Figure 39, 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 institutions into the three core business 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 asset side diversification that the *Whole-*

sale model offers to institutions that used to have 73% of their assets invested in retail loans (see Table 22). 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.

Table 31 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 geographic 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 C.3.2 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 31 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 identifies also 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 31: 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*.

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%									

As our main goal is to recognize the driving forces behind the wave of financial distress originated from the 2007-08 crisis, we circumscribe the analysis to both switches from 2005 to 2006 and from 2006 to 2007. To investigate the reasons for switching business models we propose a basic framework similar to the one applied to study the dynamics of failures. 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 consider three potential determinants: *i)* the size of the institution, to understand whether changes in business model membership are more likely for instance among big players; *ii)* macroeconomic conditions, to assess whether external factors influence the likelihood to modify the respective business model; iii) the membership to a certain business model, to understand whether being in some peer groups 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. The *GDP per Capite* is introduced to verify how economic 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 belong to in time *t*. To ameliorate potential endogeneity problems, we rely on the averages (three years before) of the economic regressors' values.

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 which are farther from the core of their group in a certain time are more likely to be candidates to switch to a different model in the next period. To measure this distance, we introduce the *Group-Score* (hereinafter *G-Score*), i.e. within each group we compute the sum of the distances between each observation from the mean of each measure 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 observation of bank *i* for measure *j*, while μ_j and σ_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 (the absence of a balance sheet item can be seen itself as a sign of business model features).

Tables 32 and 33 exhibit estimates for the transition dynamics from 2005 to 2006 and from 2006 to 2007, respectively. As expected, *G-Scores* have a significant and positive effect, either within a single group or in the whole sample. Hence, institutions that are less central in a certain group are more prone to switch into a different one, regardless the initial business model (as approximated by the peer group of origin). Less clear is the impact of *Total Assets*. One might argue that large institutions are more 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 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. From 2005 to 2006, significant estimates indicate a positive effect of *Total Assets* for *Commercial* and *Focus Retail* groups, while once we consider the entire sample estimates become not significant. By contrast, from 2006 to 2007 estimates are generally positive and significant even in the all sample. Thus, being larger in terms of total assets seems to slightly facilitate a switching dynamics at least before the outbreak of 2007-08. In few cases the impact of macro-economic conditions is significant and the sign of the coefficient is in general positive although modest (with the exception of the *Wholesale* model, from to 2006 to 2007, 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.

We also notice in Table 32 that institutions belonging to the *Long Term* model are more likely to change their group membership from 2005 to 2006 than institutions in the *Focus Retail* group, as also suggested by the transition flows shown in Figure 39. In 2007 these two models disappeared: *Long Term* institutions moved mainly to the *Commercial* group, while *Focus Retail* institutions migrated to the *Wholesale* group. Finally, Table 33 provides some insights on the impact of a previous switch. The coefficient for the dummy *Switch 2006* is positive and significant, thus having switched model between 2005 and 2006 made those institutions more unstable and more likely to redefine their business models in the next period. In addition, we provide in Appendix C.1 a robustness analysis to overcome some of the potential issues related to the presence of missing values especially for certain groups. Here, we briefly anticipate that results on those enlarged samples support on average the findings discussed in this Section.

Table 32: Transition Models from 2005 to 2006. The first five columns refer to estimates computed for each peer group separately. *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. Peer group names reported in columns names refer to the membership in 2005. Variable *Group* stands for the membership in 2005 to a specific business model (the reference level is the *Wholesale* group). Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Long Term	Focus Retail	Model 1	Model 1*	Model 2	Model 2*
Intercept	-2.976^{***}	-4.401^{***}	-4.276^{***}	-2.101^{***}	-7.534^{***}	-4.039^{***}	-4.034^{***}	-3.988^{***}	-4.037^{***}
	(0.267)	(0.208)	(0.277)	(0.370)	(1.330)	(0.123)	(0.134)	(0.151)	(0.161)
Total Assets (avg 2003-05)	-2.199	2.378^{**}	-17.794	-0.331	13.201^{**}	-0.095	-0.075	-0.272	-0.072
	(2.190)	(1.189)	(11.569)	(1.540)	(6.559)	(0.758)	(0.850)	(0.768)	(0.850)
GDP per capita (avg 2003-05)	-0.025	0.134^{**}	-0.010	0.048	0.190	0.085^{***}	0.038	0.045^{**}	0.038
	(0.043)	(0.061)	(0.050)	(0.042)	(0.210)	(0.021)	(0.026)	(0.023)	(0.027)
G-Score (2005)	0.864^{***}	2.211^{***}	2.570^{***}	0.898^{**}	3.509^{***}	1.914^{***}	1.973^{***}	1.888^{***}	1.972^{***}
	(0.283)	(0.212)	(0.305)	(0.357)	(1.120)	(0.127)	(0.138)	(0.128)	(0.138)
Group(2005): Commercial								0.013	-0.006
								(0.138)	(0.140)
Group(2005): Saving								0.008	0.011
								(0.138)	(0.138)
Group(2005): Long Term								1.055^{***}	
								(0.167)	
Group(2005): Focus Retail								-1.507^{***}	
								(0.420)	
Num. obs.	1591	1858	1536	344	338	5667	4985	5667	4985
Num. Switch	125	128	115	80	6	454	368	454	368
McFadden's Pseudo R^2	0.012	0.153	0.089	0.022	0.359	0.077	0.074	0.094	0.070
McFadden's Adjusted Pseudo R^2	0.003	0.142	0.076	0.000	0.186	0.074	0.070	0.088	0.065
LR Test	10.296^{**}	156.052^{***}	84.006^{***}	8.078^{**}	26.488^{***}	262.271^{***}	212.936^{***}	331.237^{***}	212.863^{***}

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 33: Transition Models from 2006 to 2007. The first three columns refer to estimates computed for each peer group separately. *Model 1* includes 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. Peer group names reported in columns names refer to the membership in 2006. Variable *Group* stands for the membership in 2006 to a specific business model (the reference level is the *Wholesale* group). Dummy *Switch 2006* refers to a switch from period 2005 to 2006. Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Model 1	Model 1*	Model 2*	Model 3*
Intercept	-1.837^{***}	-3.749^{***}	-3.819^{***}	-2.236^{***}	-3.324^{***}	-3.258^{***}	-3.188^{***}
	(0.242)	(0.194)	(0.231)	(0.081)	(0.120)	(0.143)	(0.145)
Total Assets (avg 2004-06)	0.677	1.858^{**}	-0.436	0.842^{***}	0.926^{**}	0.883^{**}	0.947^{**}
	(0.512)	(0.846)	(1.593)	(0.285)	(0.404)	(0.407)	(0.416)
GDP per capita (avg 2004-06)	-0.163^{***}	0.062	0.203^{***}	0.044^{***}	0.019	0.012	-0.007
	(0.035)	(0.053)	(0.039)	(0.013)	(0.021)	(0.021)	(0.022)
G-Score (2006)	0.078	1.428^{***}	1.194^{***}	1.071^{***}	1.192^{***}	1.198^{***}	0.925^{***}
	(0.271)	(0.182)	(0.264)	(0.092)	(0.121)	(0.123)	(0.127)
Group(2006): Commercial						-0.178	-0.177
						(0.124)	(0.124)
Group(2006): Saving						-0.33	-0.022
						(0.118)	(0.120)
Switch 2006							1.531^{***}
							(0.128)
Num. obs.	2234	1916	1620	6674	5770	5770	5770
Num. Switch	195	136	143	1378	474	474	474
McFadden's Pseudo R^2	0.008	0.065	0.053	0.020	0.023	0.021	0.057
McFadden's Adjusted Pseudo R^2	0.000	0.055	0.043	0.019	0.020	0.016	0.052
LR Test	29.274^{***}	78.287^{***}	66.568^{***}	163.396^{***}	96.692^{***}	99.858^{***}	222.828^{***}

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

4.7 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 distress events (presented in Table 18). 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 4.3.3 (see Table 19). Due to the nature of these regressors, we might suspect the presence of potential multicollinearity issues. Table 35 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 model for risk assessment, where we basically select subsets of regressors to overcome potential issues related to multicollinearity. Summary statistics of regressors' distributions are provided in Table 34.

Table 34: Regressors Summary Statistics. Values are in percentage and are computed as averages over the interval 2005-07. For variables definitions see Section 4.3.3.

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

Table 35: 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.

	Capital	Cap. Fund. Ratio	Roa	Cost to Inc. Ratio	Roe	Net Int. Margin	Int. Exp. to Tot Liab.	Liq. Assets to ST Fund.	Dep. to Tot Fund.	Tot Sec. to Tot Assets	GDP per Capita	Inflation	House Price	Unempl.	FDI-In.	FDI-Out.	Cent. Gvt. Debt	Gvt. LT Yield	Bank NPLs to Gross Loans	Credit to Non-Fin. Sector	Mkt Index	Sector Index	Stock Traded	Total Assets
Capital		0.9513*	0.4743*	−0.0482*	0.0444*	0.2249*	0.0893*	0.4309*	−0.0970*	0.0076	0.2195*	0.3188*	0.3458*	0.0210	0.0492*	−0.0315*	−0.1514*	0.2813*	0.0502*	−0.2453*	0.2024*	0.3127*	0.0171	−0.0813*
Capital Funding Ratio	0.9513*		0.4387*	−0.0538*	0.0369*	0.2224*	0.0915*	0.4200*	−0.1077*	−0.0089	0.2240*	0.3179*	0.3429*	0.0232	0.0502*	−0.0299*	−0.1761*	0.2936*	0.0377*	−0.2480*	0.2050*	0.3048*	0.0255	−0.0720*
Roa	0.4745*	0.4390*		−0.3432*	0.5950*	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*	−0.0364*	−0.3540*		−0.4316*	−0.0187	−0.0543*	0.0180	0.1203*	0.0194	−0.0510*	−0.0524*	−0.0740*	0.0574*	−0.1088*	−0.0634*	0.0510*	−0.0474*	−0.0129	−0.0052	−0.0470*	−0.0309	−0.0208	−0.0295
Roe	0.0446*	0.0373*	0.5715*	−0.4223*		0.1397*	0.0073	0.0694*	−0.0392*	−0.0293	0.1449*	0.2293*	0.2230*	−0.0090	0.1138*	0.0494*	−0.1956*	0.2343*	−0.0128	−0.1559*	0.1321*	0.1508*	0.0130	0.0470*
Net Interest Margin	0.2254*	0.2227*	0.2293*	−0.0194	0.1393*		0.0282	0.0496*	−0.0365*	−0.1416*	0.2212*	0.3215*	0.2481*	0.0960*	−0.0326*	−0.0691*	−0.1445*	0.2919*	0.0740*	−0.2661*	0.2415*	0.2802*	−0.0686*	−0.0596*
Interest Expenses to Total Liabilities	0.0886*	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	−0.0403*	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	0.1532*	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	−0.0979*	−0.1087*	−0.0651*	0.1156*	−0.0334*	−0.0322*	−0.0175	−0.2896*		0.0494*	−0.0097	−0.1096*	−0.2345*	0.0726*	−0.0314*	0.0206	−0.0245	−0.1166*	−0.0722*	0.1653*	0.0044	−0.1568*	−0.0098	−0.1416*
Total Securities to Total Assets	0.0076	−0.0127	0.0331*	0.0094	−0.0396*	−0.1435*	−0.0017	0.0380*	0.0495*		−0.1877*	−0.1761*	−0.2563*	−0.0084	−0.0536*	−0.0112	0.1861*	−0.1258*	−0.0113	−0.0038	−0.1461*	−0.2313*	−0.0205	0.1448*
GDP per Capita	0.2199*	0.2240*	0.1482*	−0.0503*	0.1430*	0.2204*	0.0256	0.1385*	−0.0095	−0.1887*		0.7192*	0.7123*	0.2048*	0.2080*	0.0379*	−0.5484*	0.5556*	0.0896*	−0.3251*	0.7596*	0.8262*	−0.2647*	−0.0425*
Inflation	0.3189*	0.3180*	0.2555*	−0.0496*	0.2237*	0.3224*	0.0548*	0.1644*	−0.1080*	−0.1777*	0.7167*		0.8147*	0.1033*	0.0093	−0.0577*	−0.5259*	0.8589*	0.1870*	−0.6856*	0.6404*	0.8549*	−0.0980*	−0.0527*
House Price	0.3458*	0.3429*	0.2462*	−0.0726*	0.2159*	0.2481*	0.0271	0.2100*	−0.2341*	−0.2566*	0.7133*	0.8147*		−0.0429*	0.0967*	−0.0099	−0.4136*	0.6121*	−0.0151	−0.4333*	0.6176*	0.9031*	−0.0278	−0.0123
Unemployment	0.0210	0.0232	−0.0278	0.0584*	−0.0075	0.0965*	0.0088	−0.0084	0.0728*	−0.0111	0.2048*	0.0999*	−0.0429*		−0.0775*	−0.1053*	−0.2613*	0.1332*	0.3113*	−0.0429*	0.4127*	0.0811*	−0.4928*	−0.0468*
FDI-Inflows	0.0300*	0.0364*	0.0373*	−0.0655*	0.0580*	−0.0211	0.0045	0.0677*	−0.0444*	−0.0311*	0.2093*	0.0097	0.0967*	−0.0738*		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	−0.0104	0.0361*	−0.0590*	−0.0099	−0.1040*	0.2520*		−0.1790*	−0.1107*	−0.1028*	0.0539*	−0.0548*	−0.0502*	−0.1292*	0.0085
Central Gvt. Debt	−0.1552*	−0.1790*	−0.1692*	0.0544*	−0.1893*	−0.1455*	−0.0401*	−0.0643*	−0.0233	0.1828*	−0.5362*	−0.5191*	−0.4134*	−0.2531*	−0.1992*	−0.1778*		−0.5727*	0.2052*	0.1360*	−0.6113*	−0.4035*	−0.0993*	0.0275
Gvt. Long-Term Yield	0.2813*	0.2936*	0.2279*	−0.0425*	0.2265*	0.2919*	0.0562*	0.0984*	−0.1159*	−0.1260*	0.5556*	0.8589*	0.6121*	0.1332*	−0.0186	−0.1107*	−0.5727*		0.0847*	−0.7421*	0.5556*	0.6169*	0.0665*	−0.0280
Bank NPLs to Gross Loans	0.0536*	0.0401*	0.0149	−0.0111	−0.0102	0.0841*	0.0125	0.0081	−0.0656*	−0.0168	0.1055*	0.2038*	−0.0160	0.3166*	−0.0426*	−0.1055*	0.1967*	0.0884*		−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*	−0.0047	−0.3251*	−0.6856*	−0.4333*	−0.0429*	0.2466*	0.0539*	0.1360*	−0.7421*	−0.1689*		−0.2556*	−0.4472*	−0.0624*	0.0372*
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*	0.0696*	0.0647*	−0.5903*	0.5441*	0.3126*	−0.2569*		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*	0.8549*	0.9031*	0.0811*	0.0446*	−0.0510*	−0.4022*	0.6169*	0.1095*	−0.4472*	0.7698*		−0.6289*	−0.0517*
Stock Traded	0.0094	0.0179	0.0145	−0.0163	0.0043	−0.0779*	0.0033	−0.0709*	−0.0112	−0.0182	−0.2713*	−0.1218*	−0.0359*	−0.4795*	−0.0450*	−0.1229*	−0.0890*	0.0370*	−0.4363*	−0.0621*	−0.5797*	−0.6282*		0.0280
Total Assets	−0.0813*	−0.0720*	−0.0264	−0.0284	0.0448*	−0.0599*	−0.0042	0.0479*	−0.1410*	0.1451*	−0.0425*	−0.0528*	−0.0123	−0.0468*	−0.0335*	0.0087	0.0279	−0.0280	−0.0439*	0.0372*	−0.0635*	−0.0517*	0.0299	

We first focus on the association between distress events and business models. Table 36 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[26] in the sample should not be taken lightly as they account for a large proportion of total assets (see e.g. Figure 40). For *Wholesale* group, the 64 banks under distress account of up to \$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 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 difficult. 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.

[26]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).

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 40: 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 refers to the entire interval 2005-07 (period 2005-06 for Long Term and Focus Retail) to that group and is emphasized by continuous lines.

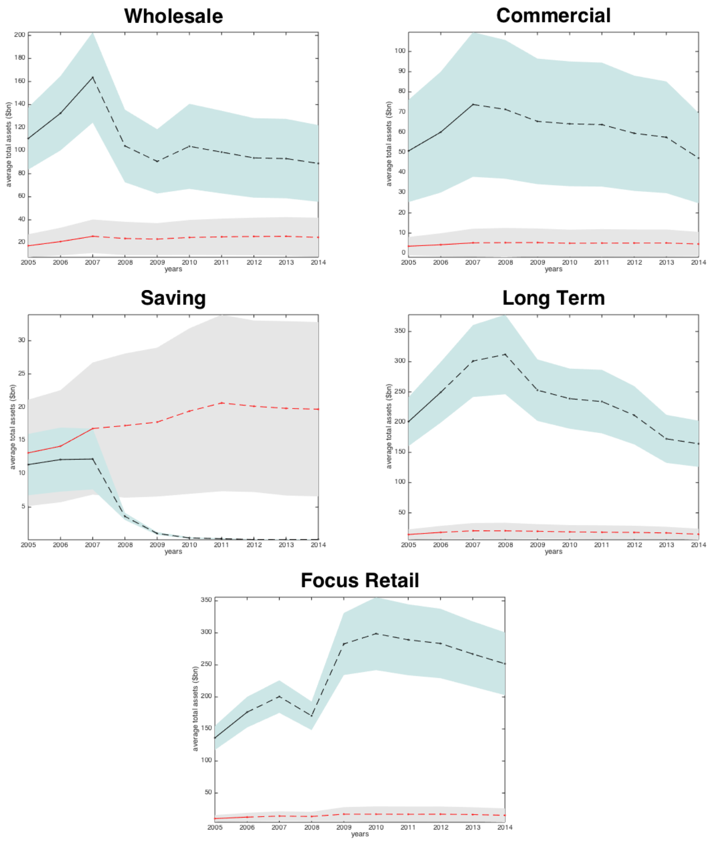


Table 36: Distribution of Distress Events for Model Membership. Column names refer to peer groups in year 2007; row names in bold stand for peer groups in year 2006; finally, row names in italics refer to peer groups in 2005. Each cell represents the number of distressed institutions referring to the 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 we need to sum up across groups in 2006.

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

Table 37 shows the results of the first model specifications where we introduce the main reference 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 variables, respectively. As reported by Betz et al. (2014) and in the references therein, CAMELS dimensions provide a good representation of the resilience of an institution. Table 37 indicates that negative and significant effects are associated to capital, 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. Instead, more liquid positions influence this risk negatively, as higher levels of liquidity might enforce the solvency of the institution. Less clear is the impact of capital, as more equity seems to favour safer levels, although once considered together with subordinated debts the sign becomes positive. 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. 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. Odd sign of unemployment in the Macro model disappears when we introduce a more comprehensive list of regressors. Expected 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 35. Marginal negative effects appear for the central government debt exposure and even for the proportion of domestic financial sector NPLs to total loans. Opposite contribution to institutions' resilience is associated with sovereign debt yields 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[27]. Therefore, financial institutions are used to present relevant exposures to sovereign debt, especially domestic issuances, which influence banking practices and the likelihood of being vulnerable to distress from shocks in sovereign debt markets (see e.g. Laeven and Valencia 2012). For completeness, we also run a specification with all regressors, even though we note that the large number of variables compared with the number 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 we notice that ROA and liquidity (deposits) 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 (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. The last model specification (*Benchmark*) takes also into account membership volatility of institutions in switching business models before the onset of the crisis[28]. We aim at testing

[27] For instance prior to the crisis banks under Basel regulations usually benefit from lower risk-weigths for positions on government bonds instead of loans. For details see e.g. BCBS 2013.

[28] 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. The resulting institutions that are discarded because they belong to vary 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 available.

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. Note that institutions in the *Long Term* and *Focus Retail* groups all moved to one of the three core peer groups, which does not necessarily mean that they have substantially modified their business models. Hence, *two* switches in the three year period before 2008 identify those institutions that present really unstable peer group membership and that represent a better proxy for the volatility of business model membership.

Table 37: 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 parsimoniuos 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 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 4.3.3. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimensions.

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

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

As we have seen in Table 37, an erratic migration dynamics can favour the probability of distress events. Thus, next analysis focuses specifically on the switching behaviour and introduces interaction effects to estimate whether belonging to certain business models or changing peer group membership affect the risk of distress. Once again we rely on a Penalized-likelihood logistic regression, 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[29]. 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 37. 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 39). Estimates in Table 38 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 37. 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, respectively. 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. Also, we note that the coefficient for the switching categorical variable (=2 switches) is significant and positive as we have previously seen in Table 37, and this is a further confirmation that this very parsimonious framework provides

[29]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; in addition, we exploit additional data from FDIC for distressed institutions. In particular, regressors for the Commercial group present several missing values, which heavily reduce the number of distressed institutions once we use complete cases without proxies in the regression.

estimates similar to those obtained with a more structured setup. In addition, the presence of significant estimates for *Switch Group* is also interesting as motivating a deeper investigation of the actual flows across business models. Hence, *Model F* exhibits estimates where we decompose the switches into the different combinations of migrations across peer groups between years 2006 and 2007. 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 4.5.2, the latter transition from a wholesale-oriented to a deposit-oriented model might be a sign of funding constraints (limited availability of debt funding due to increased perception of risk of the institution that could have caused a liquidity shortfall and a sudden reduction of wholesale funding compared to the deposits). Furthermore, transitions into the *Wholesale* group determine a worsening in 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. 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 or not (see e.g. Ayadi et al. 2012; Roengpitya et al. 2014). Estimates point to the absence of significant effects for both business models membership and for transitions across them, which seems to support the intuition that traditional broad classifications may be less likely to detect relevant effects and to differentiate the risk among business models. This is an important contribution of the study, since the introduction of a more detailed representation of banking activities allows us to better discriminate among institutions and to place them into a more accurate and suitable peer group.

Table 38: 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 collapsed 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 4.3.3. Due to the presence of missing values which affect the coverage for some groups, thus preventing a proper study of the interaction effects across them, we prefer to add some geographical aggregated proxies for macro and sectoral variables. We also admit the presence of one missing value in the computation of average values for regressors and, in addition, we exploit additional data from FDIC for distressed institutions. Superscripts *C*, *A*, *M*, *E*, *L*, *S* indicate the respective CAMELS dimension.

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Intercept	-0.929^{**}	-0.996^{**}	-1.131^{***}	-0.759	-0.665	-0.747	-1.404^{***}
	(0.463)	(0.454)	(0.423)	(0.464)	(0.473)	(0.469)	(0.449)
$^{C}Capital$	-0.001	0.001	-0.004	0.000	0.001	0.001	-0.003
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
^{A}Roa	-0.258^{***}	-0.257^{***}	-0.271^{***}	-0.261^{***}	-0.258^{***}	-0.262^{***}	-0.273^{***}
	(0.070)	(0.070)	(0.071)	(0.069)	(0.069)	(0.067)	(0.072)
^{M}Roe	0.042^{***}	0.042^{***}	0.039^{***}	0.042^{***}	0.042^{***}	0.041^{***}	0.039^{***}
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
$^{E}Net\ Interest\ Margin$	-0.056	-0.061	-0.074^{*}	-0.057	-0.055	-0.037	-0.083^{**}
	(0.043)	(0.043)	(0.039)	(0.042)	(0.042)	(0.041)	(0.037)
$^{L}Deposits\ to\ Total\ Funding$	-0.022^{***}	-0.022^{***}	-0.019^{***}	-0.023^{***}	-0.023^{***}	-0.023^{***}	-0.018^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$^{S}Total\ Securities\ to\ Total\ Assets$	0.004	0.004	0.003	0.003	0.003	0.003	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
GDP per capita	-0.201^{**}	-0.173^{**}	-0.132^{*}	-0.201^{**}	-0.206^{**}	-0.260^{***}	-0.104
	(0.088)	(0.085)	(0.077)	(0.087)	(0.087)	(0.088)	(0.074)
Inflation	0.445^{***}	0.425^{***}	0.353^{***}	0.437^{***}	0.442^{***}	0.484^{***}	0.353^{***}
	(0.098)	(0.094)	(0.085)	(0.096)	(0.097)	(0.098)	(0.083)
Gvt. Long-Term Yield	-0.081	-0.073	-0.022	-0.086	-0.095	-0.117	-0.010
	(0.093)	(0.091)	(0.080)	(0.093)	(0.095)	(0.095)	(0.077)
Market Index	-0.078^{***}	-0.079^{***}	-0.071^{***}	-0.078^{***}	-0.079^{***}	-0.077^{***}	-0.073^{***}
	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)

Distress Assessment: Transition Analysis: *continued*

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
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^{***}		1.379^{***}	1.382^{***}		
		(0.252)		(0.390)	(0.398)		
Group (2006): Saving		0.019		0.283	0.129		
		(0.216)		(0.408)	(0.443)		
Group (2006): Long Term		-0.654^{***}		-0.014	0.129		
		(0.252)		(0.352)	(0.427)		
Group (2006): Focus Retail		0.534		0.470	0.708		
		(0.353)		(0.354)	(0.434)		
Group (2007): Commercial			-0.384^{*}	-0.861^{**}	-0.817^{**}	0.871^{*}	
			(0.199)	(0.354)	(0.358)	(0.467)	
Group (2007): Saving			-0.102	-0.315	-0.170	-0.773	
			(0.202)	(0.396)	(0.430)	(0.680)	
Group (2007): Wholesale-oriented							-0.409
				(0.396)			(0.852)
Switch Group = 1					-0.297		
					(0.286)		
Switch Group = 2					0.962^{*}		
					(0.567)		
Group (2006) Commercial $\longmapsto$ Group (2007) Wholesale						1.884^{***}	
						(0.634)	
Group (2006) Saving $\longmapsto$ Group (2007) Wholesale						0.746	
						(0.884)	
Group (2006) Long Term $\longmapsto$ Group (2007) Wholesale						0.689^{*}	
						(0.404)	
Group (2006) Focus Retail $\longmapsto$ Group (2007) Wholesale						0.507	
						(0.367)	
Group (2006) Commercial $\longmapsto$ Group (2007) Commercial						-0.224	
						(0.494)	
Group (2006) Saving $\longmapsto$ Group (2007) Commercial						-1.747	
						(1.518)	
Group (2006) Long Term $\longmapsto$ Group (2007) Commercial						-2.104^{***}	
						(0.517)	
Group (2006) Focus Retail $\longmapsto$ Group (2007) Commercial						1.441	
						(1.134)	
Group (2006) Commercial $\longmapsto$ Group (2007) Saving						0.123	
						(1.688)	
Group (2006) Saving $\longmapsto$ Group (2007) Saving						0.788	
						(0.678)	
Group (2006) Long Term $\longmapsto$ Group (2007) Saving						2.351^{***}	
						(0.874)	
Group (2006) Wholesale-oriented $\longmapsto$ Group (2007) Saving							-0.124
							(0.438)
Group (2006) Wholesale-oriented $\longmapsto$ 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^2	0.111	0.108	0.101	0.109	0.110	0.122	0.099
McFadden's Adjusted Pseudo R^2	0.092	0.089	0.084	0.087	0.085	0.092	0.080

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

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, we analyse each business model separately by partitioning the sample according to stable membership along the interval 2005-07. Table 39 focuses on a restricted case where we discard institutions that switched business models prior to the crisis[30]. 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. However, estimates for small groups should be taken with care due to data limitations. 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 stand for market dynamics (Government LT Yield and Market Index return). The choice of these regressors is in line to the ones presented in Table 38 producing estimates for the entire set of no-switching institutions similar to those discussed in the Benchmark model of Table 37. This supports the selection of these regressors for the reference model used to create specifications for each business model, which although confined to specific peer groups membership though share the same setup to enhance comparability. In addition, we add the amount of institutions' total assets as regressor. In fact, as highlighted in Section 4.5 and in the discussion therein for the single peer groups, relative size seems to matter for the event of distress during the recent financial crisis at least for the wholesale-oriented institutions. As a result, dominant relative asset size within the peer group 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 also in terms of CAMELS variables: ROE impacts positively on the likelihood of distress for *Wholesale* institutions and nega-

[30]For Long Term and Focus Retail groups we consider stable participation of the institutions in these groups during the biennium 2005-06.

tively 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 emphasize different forces affecting institutions' probability of distress under wholesale-oriented or deposit-oriented models. We might advance the explanation that institutions within the *Saving* model cannot fully exploit a wide spectrum of investment choices to boost their returns on assets. This is due to a limited, compared to the *Wholesale*, range of available instruments in the assets side, which probably impacts on their ability to select profitable investments. However, *Saving* institutions might find convenient to enlarge their business into more profitable investments, which may favour more diversification on the assets side and, possibly, facilitate even more resilience. Conversely, institutions adopting the *Wholesale* model may already present higher exposure on profitable and risky assets, and further investment decisions to increase the level of ROE can worsen the sustainability of their activities. This might be due to the structure of the liabilities side: *Wholesale* institutions have higher levels of debts, especially interbank debts, and further investments are indeed likely to involve an increase on leverage, which in turn deteriorates the resilience of their business model and, eventually, exacerbates their risk of distress; *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. The latter phenomenon can imply contagion dynamics that are related to interbanking exposures through which contagion may actually propagate. 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 in deposit-oriented institutions, were not particularly affected by bank runs triggered by the lack of confidence in banks quality. Furthermore, the interconnectivity arisen from interbanking 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. As mentioned earlier, the impact of total assets is positive and significant for *Wholesale*, but for the *Saving* group the coefficient is modest and not significant. This supports the discussion above on the importance of relative size in the risk assessment of distress, thus remarking that big institutions in terms of total assets belonging to wholesale-oriented models are more likely to have suffered from distress during the recent crisis, while distresses for those with *Saving* business model seem to be not particularly affected by institutions' sizes.

We observe that *Commercial* group has few distress events mainly concentrated in US (22 out of 26 in the model presented in Table 39), that might explain why macro and sectoral dimensions play a significant role for them. This business model has also a positive and significant coefficient for the net interest margin, differently from the other core models presented in Table 39. Due to few observations of distress events for this peer group and the lack of appropriate coverage of balance sheet items provided by Bankscope, here we enrich the dataset with information collected from FDIC on distressed institutions; in addition, we admit the presence of one potential missing value in the computation of average values for the regressors and we also replace missing values for macro and sectoral variables with geographical aggregates. This approach enlarges the coverage of the sample and allows us to reasonably estimate the model. In Appendix C.1 we also present estimates for the other peer groups using similar proxies; here, we anticipate that the overall picture is confirmed under the approximated scenario even for the other peer groups' specifications. Although we are aware that a proper econometric analysis for the other two marginal groups should probably consider a smaller list of regressors, we still estimate the model for these two groups to provide comparisons with the three core business models. *Long Term* group has negative and significant coefficients for ROA and 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. Finally, we emphasize that capital has a significant and negative sign only for the *Saving* model thus supporting the presence of an overconfidence in the importance of capital levels for bank risk assessment at least for certain business models (a result in part already presented in Vazquez and Federico 2015). Furthermore, similar to the previous case shown in Table 38 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). Estimates reinforce the interpretation that wholesale-oriented and deposit-oriented models present different drivers for the risk of distress.

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 addition 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 with a different business model. 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. A 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 39: Distress Assessment within Peer Groups. Column *All* includes only institutions that do not switch peer groups in the interval 2005-07 (in the case of Long Term and Focus Retail the interval that is considered is 2005-06). The other columns 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. Column *Wholesale-oriented (Restricted)* is circumscribed to Wholesale and Commercial models. For regressors definitions see Section 4.3.3. 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	-2.071^{***}	-1.130	-0.968	4.942^{***}	-3.770^{***}	-5.490^{*}	-2.564^{***}	-1.815^{**}
	(0.497)	(1.108)	(1.956)	(1.764)	(1.385)	(2.866)	(0.628)	(0.925)
$^{C}Capital$	-0.010	0.010	0.034	-0.138^{**}	-0.001	-0.234	0.006	0.018
	(0.014)	(0.013)	(0.050)	(0.055)	(0.064)	(0.264)	(0.013)	(0.012)
^{A}Roa	-0.163	-0.175^{*}	-0.380	0.678^{**}	-0.996^{***}	1.065	-0.240^{**}	-0.252^{**}
	(0.100)	(0.094)	(0.249)	(0.286)	(0.364)	(2.176)	(0.104)	(0.105)
^{M}Roe	0.026^{***}	0.044^{***}	0.052	-0.042^{**}	0.007	0.028	0.038^{***}	0.061^{***}
	(0.009)	(0.015)	(0.040)	(0.018)	(0.021)	(0.128)	(0.011)	(0.015)
$^{E}Net\ Interest\ Margin$	-0.093^{**}	0.012	0.147^{***}	0.169	-0.589^{**}	-0.040	-0.082^{**}	-0.025
	(0.036)	(0.009)	(0.056)	(0.166)	(0.280)	(0.451)	(0.036)	(0.049)
$^{L}Deposits\ to\ Total\ Funding$	-0.017^{***}	-0.020^{***}	0.024	-0.066^{***}	0.004	-0.015	-0.017^{***}	-0.016^{***}
	(0.004)	(0.006)	(0.020)	(0.015)	(0.012)	(0.015)	(0.005)	(0.006)
$^{S}Total\ Securities\ to\ Total\ Assets$	-0.002	0.004	0.022	-0.003	0.019	-0.090	0.003	0.004
	(0.006)	(0.008)	(0.022)	(0.012)	(0.015)	(0.066)	(0.007)	(0.008)
GDP per capita	-0.026	-0.144	-0.966^{***}	-3.061^{***}	0.099	0.249	0.082	-0.043
	(0.093)	(0.206)	(0.350)	(1.048)	(0.213)	(0.571)	(0.100)	(0.171)
Inflation	0.300^{**}	0.203	1.962^{***}	-0.150	0.420	-2.220	0.141	0.319
	(0.123)	(0.220)	(0.530)	(0.486)	(0.361)	(1.644)	(0.135)	(0.196)
Gvt. Long-Term Yield	0.124	-0.275	-0.905^{***}	0.852^{*}	0.230	2.431^{**}	0.050	-0.242
	(0.128)	(0.287)	(0.305)	(0.483)	(0.411)	(1.146)	(0.170)	(0.249)
Market Index	-0.079^{***}	-0.021	-0.308^{***}	0.088	-0.050	-0.060	-0.050^{***}	-0.057^{***}
	(0.011)	(0.022)	(0.057)	(0.062)	(0.039)	(0.089)	(0.013)	(0.017)
Total Assets	1.790^{***}	1.606^{***}	3.141^{***}	0.029	2.531^{***}	10.086	1.984^{***}	1.531^{***}
	(0.287)	(0.476)	(1.197)	(1.450)	(0.864)	(6.424)	(0.317)	(0.395)
Num. obs.	6543	1867	1458	1231	907	349	4870	3450
Num. Distress Events	147	46	26	46	18	12	92	59
McFadden's Pseudo R^2	0.111	0.137	0.743	0.180	0.208	0.333	0.107	0.093
McFadden's Adjusted Pseudo R^2	0.092	0.067	0.645	0.111	0.052	0.066	0.077	0.048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

4.8 Conclusions

The study investigates the impact of the true banking business model on vulnerability and likelihood of distress. We employ one of the largest sample of banks globally distributed that includes more than 11,000 institutions representing more than 180 countries. The investigation of true banking business models is a very challenging task, which involves the categorization of a large sample of institutions, with a vast and complex set of balance sheets quite heterogeneous in terms of their direct specializations and country of origin, into few well-defined peer groups where members run the same business model.

To overcome these issues, we use the cosine similarity measure to compare pairs of institutions and we propose 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. This approach is specifically designed to overcome limitations due to large, complex and sparse datasets, preserving a superior 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 is characterized by dominant 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 notice that *Wholesale* and *Commercial* peer groups are the only two wholesale-oriented models available throughout the entire period 2005-14, whereas the *Long Term* and *Focus Retail* were present only for the first two years and disappeared thereafter (we observe institutions adopting these models would switch mainly to the other two models in 2007 at the onset of the crisis). 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 nonincome 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.

Then, we focused on the pre financial crisis period (2005-07) to test how the additional information of the true business model specification can help the risk assessment on the likelihood of distress events occurring in the interval 2008-10. In this respect, we develop a state-of-the-art 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. This result suggests that a deeper investigation within each peer group is needed to disentangle the impact of being in a particular group on the likelihood of distress. 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 also require deeper investigation.

Due to the scarcity of distress events compared to the overall sample size, we employed a Firth's Penalized-likelihood logistic regression model to assess the contribution of institutions' characteristics, captured by both balance sheet proxies for CAMELS, their membership to certain business models and their stability over time, and controlling for both macro and sectoral indicators. 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 busting 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 those 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 non-deposit funding, were exposed to more risk of distress.

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 positive impact of ROE for *Wholesale* and negative for *Saving*; opposite effects for ROA; 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 can be dramatically misleading compared to a targeted and bespoke approach designed for business models true classifications. Regulator 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 us distance to defauls (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.

Appendix C

C.1 Robustness Checks

As a control for the switching process, we run a similar analysis considering a three-year window for the computation of the averages of regressors' values and allowing for the presence of one missing value. This enlarges the sample and partially circumscribes the problems arising from data availability related to the coverage of our dataset for the initial period (for instance balance sheet data for many institutions that are in the *Long Term* group in 2005 are not available for previous years, thus reducing significantly the number of observations for this group). Estimates (shown in Table 40) for 2005-06 confirm the relevance of the *G-Scores* as a main driver in the switching process, while the impact of size and economic conditions become less significant. The 2006-07 results are all unchanged unless the sign of the *GDP per capita* in Model 1 which becomes negative although still modest (Table 41).

Table 40: Transition Models from 2005 to 2006 with Proxies. The first five columns refer to estimates computed for each peer group separately. *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. Peer group names reported in columns names refer to the membership in 2005. Variable *Group* stands for the membership in 2005 to a specific business model (the reference level is the *Wholesale* group). Regressors values are computed over the three year before the event, admitting for the presence of one missing value to enlarge the dataset. Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Long Term	Focus Retail	Model 1	Model 1*	Model 2	Model 2*
Intercept	-2.541^{***}	-4.340^{***}	-4.105^{***}	-1.899^{***}	-6.732^{***}	-3.812^{***}	-3.828^{***}	-3.882^{***}	-3.901^{***}
	(0.242)	(0.195)	(0.255)	(0.305)	(0.950)	(0.111)	(0.123)	(0.138)	(0.149)
Total Assets (avg 2003-05)	0.551	1.806^{**}	-17.266	-0.599	3.678	0.477	0.627	0.311	0.647
	(0.649)	(0.896)	(10.748)	(1.216)	(2.534)	(0.472)	(0.512)	(0.492)	(0.511)
GDP per capita (avg 2003-05)	-0.159^{***}	0.159^{***}	0.011	0.031	-0.060	0.025	-0.022	-0.004	-0.021
	(0.038)	(0.056)	(0.046)	(0.036)	(0.181)	(0.019)	(0.023)	(0.020)	(0.025)
G-Score (2005)	0.624^{**}	2.088^{***}	2.398^{***}	0.770^{**}	3.877^{***}	1.833^{***}	1.874^{***}	1.794^{***}	1.877^{***}
	(0.261)	(0.193)	(0.279)	(0.312)	(0.836)	(0.113)	(0.124)	(0.114)	(0.124)
Group (2005): Commercial								0.066	0.040
								(0.129)	(0.131)
Group (2005): Saving								0.200	0.195
								(0.125)	(0.125)
Group (2005): Long Term								1.242^{***}	
								(0.141)	
Group (2005): Focus Retail								-0.910^{***}	
								(0.306)	
Num. obs.	2234	1985	1625	473	399	6716	5844	6716	5844
Num. Switch	158	144	136	110	12	560	438	560	438
McFadden's Pseudo R^2	0.011	0.155	0.082	0.015	0.243	0.065	0.067	0.086	0.065
McFadden's Adjusted Pseudo R^2	0.002	0.145	0.071	-0.001	0.145	0.063	0.064	0.081	0.060
LR Test	30.743^{***}	173.604^{***}	88.946^{***}	7.633^{*}	33.049^{***}	272.994^{***}	229.507^{***}	369.759^{***}	232.081^{***}

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 41: Transition Models from 2006 to 2007. The first three columns refer to estimates computed for each peer group separately. *Model 1* includes 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 during the biennium 2005-07. Peer group names reported in columns names refer to the membership in 2006. Variable *Group* stands for the membership in 2006 to a specific business model (the reference level is the *Wholesale* group). Dummy *Switch 2006* assumes value 1 if there is an institution switches its peer group from period 2005 to 2006. Regressors values are computed over the three year before the event, admitting for the presence of one missing value to enlarge the dataset. Total Assets are in USD Trillion.

	Wholesale	Commercial	Saving	Model 1	Model 1*	Model 2*	Model 3*
Intercept	-1.756^{***}	-3.639^{***}	-3.708^{***}	-1.071^{***}	-3.217^{***}	-3.216^{***}	-3.175^{***}
	(0.221)	(0.171)	(0.212)	(0.064)	(0.110)	(0.132)	(0.133)
Total Assets (avg 2004-06)	0.853	1.602^{**}	-0.433	0.618^{**}	1.001^{**}	0.982^{**}	1.030^{**}
	(0.525)	(0.779)	(1.577)	(0.275)	(0.408)	(0.410)	(0.419)
GDP per capita (avg 2004-06)	-0.190^{***}	0.087^{*}	0.172^{***}	-0.099^{***}	0.004	0.001	-0.011
	(0.032)	(0.047)	(0.033)	(0.012)	(0.019)	(0.020)	(0.020)
G-Score (2006)	0.059	1.297^{***}	1.270^{***}	0.536^{***}	1.141^{***}	1.140^{***}	0.916^{***}
	(0.247)	(0.170)	(0.237)	(0.078)	(0.111)	(0.112)	(0.117)
Group(2006): Commercial						-0.092	-0.112
						(0.111)	(0.112)
Group(2006): Saving						0.149	0.072
						(0.109)	(0.111)
Switch 2006							1.384^{***}
							(0.117)
Num. obs.	2711	2334	1732	8507	6777	6777	6777
Num. Switch	229	174	167	2300	570	570	570
McFadden's Pseudo R^2	0.018	0.050	0.055	0.009	0.020	0.019	0.049
McFadden's Adjusted Pseudo R^2	0.011	0.042	0.045	0.008	0.018	0.015	0.045
LR Test	47.012^{***}	77.232^{***}	75.464^{***}	115.002^{***}	101.887^{***}	106.329^{***}	228.023^{***}

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

To investigate the risk of distress across different business models we present in Table 42 the same models as shown in Table 39 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; in addition, we also exploit additional data from FDIC for distressed institutions. Results are in line with those shown in Table 39, with estimates slightly more significant for macro and sectoral regressors and less for CAMELS.

Table 42: Distrees Assessment within Peer Groups using Proxies. Column *All** includes only institutions that do not switch peer groups in the interval 2005-07. The other columns 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. Column *Wholesale-oriented (Restricted)* is circumscribed to Wholesale and Commercial models. For regressors definitions see Section 4.3.3. 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 dimension. Total Assets are in USD Trillion.

	All*	Wholesale*	Commercial*	Saving*	Long Term*	Focus Retail*	Wholesale-oriented*	Wholesale-oriented* (Restricted)
Intercept	-1.677^{***}	-0.245	-0.968	3.957^{***}	-2.560^{**}	-4.901^{*}	-2.338^{***}	-1.672^{***}
	(0.409)	(1.053)	(1.956)	(1.452)	(1.056)	(2.790)	(0.476)	(0.614)
$^{C}Capital$	0.002	0.006	0.034	-0.102^{*}	-0.003	-0.217	0.010	0.014
	(0.010)	(0.013)	(0.050)	(0.052)	(0.065)	(0.295)	(0.010)	(0.010)
^{A}Roa	-0.252^{***}	-0.129	-0.380	0.642^{**}	-0.257	-0.044	-0.274^{***}	-0.307^{***}
	(0.070)	(0.091)	(0.249)	(0.284)	(0.318)	(2.616)	(0.077)	(0.077)
^{M}Roe	0.036^{***}	0.046^{***}	0.052	-0.040^{**}	0.000	0.111	0.046^{***}	0.065^{***}
	(0.008)	(0.015)	(0.040)	(0.018)	(0.023)	(0.137)	(0.010)	(0.013)
$^{E}Net\ Interest\ Margin$	-0.060	-0.027	0.147^{***}	0.047	-0.791^{***}	0.168	-0.049	0.013
	(0.042)	(0.045)	(0.056)	(0.168)	(0.239)	(0.427)	(0.046)	(0.009)
$^{L}Deposits\ to\ Total\ Funding$	-0.015^{***}	-0.020^{***}	0.024	-0.061^{***}	0.004	-0.023^{*}	-0.009^{**}	-0.015^{***}
	(0.004)	(0.006)	(0.020)	(0.015)	(0.012)	(0.014)	(0.004)	(0.005)
$^{S}Total\ Securities\ to\ Total\ Assets$	-0.003	0.002	0.022	-0.009	0.023	-0.052	0.001	-0.001
	(0.005)	(0.007)	(0.022)	(0.012)	(0.014)	(0.054)	(0.006)	(0.006)
GDP per capita	-0.120	-0.039	-0.966^{***}	-1.471^{***}	-0.093	0.063	-0.065	-0.150
	(0.075)	(0.148)	(0.350)	(0.548)	(0.210)	(0.574)	(0.083)	(0.117)
Inflation	0.373^{***}	0.345^{*}	1.962^{***}	0.494	0.845^{***}	-2.820^{*}	0.348^{***}	0.330^{**}
	(0.085)	(0.205)	(0.530)	(0.304)	(0.309)	(1.633)	(0.103)	(0.130)
Gvt. Long-Term Yield	-0.043	-0.477^{*}	-0.905^{***}	0.221	-0.091	2.705^{**}	-0.069	-0.077
	(0.081)	(0.267)	(0.305)	(0.262)	(0.190)	(1.174)	(0.099)	(0.140)
Market Index	-0.073^{***}	-0.048^{***}	-0.308^{***}	-0.028	-0.097^{***}	-0.074	-0.075^{***}	-0.075^{***}
	(0.008)	(0.012)	(0.057)	(0.034)	(0.030)	(0.094)	(0.009)	(0.010)
Total Assets	1.752^{***}	1.511^{***}	3.141^{***}	-0.021	2.432^{***}	4.044	1.897^{***}	1.407^{***}
	(0.279)	(0.465)	(1.197)	(1.458)	(0.828)	(3.100)	(0.309)	(0.377)
Num. obs.	6795	1965	1458	1255	949	350	5063	3956
Num. Distress Events	183	55	26	47	21	12	126	103
McFadden's Pseudo R^2	0.123	0.078	0.743	0.169	0.228	0.315	0.129	0.145
McFadden's Adjusted Pseudo R^2	0.107	0.024	0.645	0.100	0.091	0.050	0.107	0.117

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

C.2 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)

C.3 Non-Parametric Analyis of Peer Groups

In this Appendix we compare groups according to a set of balance sheet measures (see Appendix C.2) which we selected to study differences in 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 4.4, this implies the maximization of the *modularity* of the resulting configuration. In this Appendix we introduce an additional framework to verify economically 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: from one side they aim to support the capacity of the Louvain algorithm to group similar institutions by testing statistically for median differences in groups' balance sheet measures, and from the other side these tests on a large set of measures help the interpretation of emerging groups in terms of business models features.

Appendix C.3.1 is based on results from the non-parametric equality of medians test (Kruskal-Wallis) which we applied to verify 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 have further analysed by means of post-hoc multiple pairwise comparisons (Dunn test). In Appendix C.3.1 for each year we report[1]: 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. We use the Bonferroni correction to take into account the FWER (we also apply other approaches, not re-

[1]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.

ported, like Sidak and Holm corrections obtaining similar results). 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. We use a restrictive significance level of 1% to judge whether two groups differ or not. In addition, in order to remove some noise due to the presence of potential outliers we 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 (column *Core*). 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, in Appendix C.3.2 we discuss some examples to verify 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.

Finally, in Appendix C.3.3 we provide 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 4.3.1). 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.

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 (Volatile group not stable in time).

C.3.1 Multiple Comparisons

	Retail Loans				
2005	*# observations*	*Median*	*Mean*	*all*	*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	****	****

	Retail Loans				
2006	*# observations*	*Median*	*Mean*	*all*	*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	****	****

	Retail Loans				
2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.00	0.20	**	**
II	3861	0.00	0.01	**	**
III	2070	0.00	0.04	**	**

	Retail Loans				
2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.00	0.11	***	**
II	3003	0.00	0.02	***	***
VI	2053	0.34	0.38	***	***
III	1708	0.00	0.04	***	**

	Retail Loans				
2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.00	0.13	***	***
II	2877	0.00	0.02	**	***
VI	2195	0.28	0.32	***	***
III	2125	0.00	0.03	**	***

	Retail Loans				
2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.00	0.13	***	***
II	2760	0.00	0.02	***	***
VI	2233	0.29	0.32	***	***
III	2009	0.00	0.03	***	***

	Retail Loans				
2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.01	0.16	****	****
II	2692	0.00	0.01	****	****
VI	2252	0.30	0.34	****	****
III	1778	0.00	0.03	***	***
VIII	573	0.00	0.05	***	***

	Retail Loans				
2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.06	0.20	*****	*****
VI	2244	0.30	0.34	*****	*****
II	2125	0.00	0.01	****	***
III	1901	0.00	0.03	****	*****
VIII	1053	0.00	0.05	****	***
VII	310	0.00	0.02	****	***

	Retail Loans				
2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.01	0.13	****	****
VI	2218	0.30	0.34	****	****
II	2135	0.00	0.02	****	****
III	1849	0.00	0.04	****	****
VIII	1424	0.06	0.26	****	****

	Retail Loans				
2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.06	0.19	****	****
II	2390	0.00	0.02	***	***
VI	2086	0.32	0.35	****	****
III	1807	0.00	0.04	****	****
VII	226	0.00	0.01	***	***

		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	****	****

2005	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3221	0.57	0.51	****	****
II	2592	0.62	0.60	***	****
III	1946	0.58	0.57	****	****
IV	1416	0.65	0.61	***	****
V	606	0.81	0.78	****	****

2006	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3390	0.58	0.52	***	***
II	2568	0.63	0.62	***	***
III	1952	0.58	0.57	***	***
IV	1534	0.63	0.58	***	***
V	463	0.85	0.82	****	****

2007	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	4108	0.61	0.55	**	**
II	3861	0.65	0.64	**	**
III	2070	0.57	0.55	**	**

2008	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3270	0.58	0.50	**	**
II	3003	0.69	0.67	***	***
VI	2053	0.61	0.60	***	***
III	1708	0.58	0.56	**	**

2009	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3116	0.57	0.52	***	***
II	2877	0.66	0.61	***	***
VI	2195	0.59	0.59	**	**
III	2125	0.60	0.57	**	**

2010	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3484	0.54	0.49	***	***
II	2760	0.68	0.65	***	***
VI	2233	0.59	0.59	***	**
III	2009	0.59	0.56	***	**

2011	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3608	0.58	0.52	***	***
II	2692	0.69	0.67	****	****
VI	2252	0.61	0.61	****	****
III	1778	0.58	0.56	***	***
VIII	573	0.15	0.18	****	****

2012	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3626	0.60	0.56	****	*****
VI	2244	0.62	0.62	*****	*****
II	2125	0.69	0.67	*****	*****
III	1901	0.59	0.57	****	*****
VIII	1053	0.36	0.37	*****	*****
VII	310	0.02	0.07	*****	*****

2013	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3424	0.56	0.50	****	****
VI	2218	0.63	0.63	****	****
II	2135	0.67	0.67	****	****
III	1849	0.59	0.57	****	****
VIII	1424	0.48	0.48	****	****

2014	# observations	Median	Mean	all	core
		Retail and Corp. Loans			
I	3653	0.59	0.53	***	***
II	2390	0.64	0.62	***	***
VI	2086	0.63	0.63	***	***
III	1807	0.59	0.57	***	***
VII	226	0.01	0.07	****	****

		Total Loans			
2005	*# observations*	*Median*	*Mean*	*all*	*core*
I	3221	0.76	0.70	***	****
II	2592	0.72	0.70	***	***
III	1946	0.73	0.71	***	***
IV	1416	0.77	0.74	***	****
V	606	0.91	0.85	****	****

		Total Loans			
2006	*# observations*	*Median*	*Mean*	*all*	*core*
I	3390	0.77	0.71	***	****
II	2568	0.73	0.72	***	***
III	1952	0.73	0.72	***	***
IV	1534	0.78	0.75	***	****
V	463	0.94	0.89	****	****

		Total Loans			
2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.81	0.74	**	**
II	3861	0.76	0.75	**	**
III	2070	0.73	0.71	**	**

		Total Loans			
2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.79	0.71	**	***
II	3003	0.79	0.77	**	***
VI	2053	0.75	0.73	***	***
III	1708	0.73	0.71	***	***

		Total Loans			
2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.76	0.69	**	***
II	2877	0.77	0.73	**	***
VI	2195	0.71	0.71	**	**
III	2125	0.71	0.70	**	**

		Total Loans			
2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.74	0.67	**	***
II	2760	0.77	0.74	***	***
VI	2233	0.71	0.71	**	***
III	2009	0.70	0.69	***	***

		Total Loans			
2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.72	0.65	***	****
II	2692	0.77	0.74	****	****
VI	2252	0.73	0.73	****	****
III	1778	0.69	0.68	**	***
VIII	573	0.65	0.61	***	***

		Total Loans			
2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.73	0.69	****	****
VI	2244	0.73	0.73	***	****
II	2125	0.75	0.73	****	*****
III	1901	0.70	0.69	*****	*****
VIII	1053	0.65	0.61	*****	*****
VII	310	0.05	0.11	*****	*****

		Total Loans			
2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.70	0.63	**	**
VI	2218	0.73	0.73	***	***
II	2135	0.73	0.73	***	***
III	1849	0.70	0.69	**	**
VIII	1424	0.71	0.66	**	**

		Total Loans			
2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.72	0.67	**	**
II	2390	0.71	0.70	*	***
VI	2086	0.73	0.72	***	***
III	1807	0.70	0.69	**	****
VII	226	0.05	0.10	****	****

Interbank Lending

2005	*# observations*	*Median*	*Mean*	*all*	*core*
I	3221	0.11	0.19	****	****
II	2592	0.08	0.10	****	****
III	1946	0.10	0.14	***	***
IV	1416	0.09	0.13	***	***
V	606	0.06	0.07	****	****

Interbank Lending

2006	*# observations*	*Median*	*Mean*	*all*	*core*
I	3390	0.12	0.19	**	****
II	2568	0.08	0.10	***	****
III	1952	0.12	0.15	**	***
IV	1534	0.10	0.17	**	***
V	463	0.07	0.08	***	****

Interbank Lending

2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.11	0.20	*	*
II	3861	0.09	0.11	**	**
III	2070	0.12	0.16	*	*

Interbank Lending

2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.13	0.21	***	**
II	3003	0.07	0.09	***	***
VI	2053	0.12	0.13	**	**
III	1708	0.11	0.15	**	***

Interbank Lending

2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.12	0.19	**	***
II	2877	0.06	0.13	**	***
VI	2195	0.10	0.12	**	***
III	2125	0.08	0.13	**	***

Interbank Lending

2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.12	0.20	***	***
II	2760	0.05	0.09	***	***
VI	2233	0.09	0.12	***	***
III	2009	0.08	0.13	***	***

Interbank Lending

2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.09	0.15	***	***
II	2692	0.05	0.08	****	****
VI	2252	0.10	0.12	****	****
III	1778	0.09	0.13	***	***
VIII	573	0.45	0.44	****	****

Interbank Lending

2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.08	0.14	**	****
VI	2244	0.08	0.11	***	*****
II	2125	0.04	0.07	*****	*****
III	1901	0.07	0.13	****	****
VIII	1053	0.18	0.25	*****	*****
VII	310	0.13	0.21	***	*****

Interbank Lending

2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.09	0.16	**	****
VI	2218	0.08	0.10	**	****
II	2135	0.04	0.06	****	****
III	1849	0.07	0.13	***	****
VIII	1424	0.10	0.19	***	****

Interbank Lending

2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.08	0.15	**	***
II	2390	0.05	0.08	****	****
VI	2086	0.07	0.09	***	***
III	1807	0.06	0.12	****	****
VII	226	0.13	0.20	***	****

2005	# observations	Investments Median	Investments Mean	all	core
I	3221	0.12	0.18	***	****
II	2592	0.22	0.24	****	****
III	1946	0.20	0.21	****	****
IV	1416	0.14	0.17	***	****
V	606	0.05	0.10	****	****

2006	# observations	Investments Median	Investments Mean	all	core
I	3390	0.11	0.18	****	****
II	2568	0.21	0.22	****	****
III	1952	0.19	0.21	****	****
IV	1534	0.14	0.17	****	****
V	463	0.02	0.07	****	****

2007	# observations	Investments Median	Investments Mean	all	core
I	4108	0.09	0.16	**	**
II	3861	0.17	0.18	**	**
III	2070	0.19	0.21	**	**

2008	# observations	Investments Median	Investments Mean	all	core
I	3270	0.08	0.14	***	***
II	3003	0.13	0.15	***	***
VI	2053	0.19	0.21	**	**
III	1708	0.19	0.20	**	**

2009	# observations	Investments Median	Investments Mean	all	core
I	3116	0.10	0.16	***	***
II	2877	0.13	0.17	***	***
VI	2195	0.23	0.24	***	***
III	2125	0.19	0.20	***	***

2010	# observations	Investments Median	Investments Mean	all	core
I	3484	0.12	0.17	***	***
II	2760	0.13	0.17	***	***
VI	2233	0.24	0.24	***	***
III	2009	0.20	0.21	***	***

2011	# observations	Investments Median	Investments Mean	all	core
I	3608	0.12	0.17	**	****
II	2692	0.14	0.16	**	****
VI	2252	0.22	0.22	***	****
III	1778	0.20	0.21	***	****
VIII	573	0.11	0.17	**	****

2012	# observations	Investments Median	Investments Mean	all	core
I	3626	0.11	0.16	*****	*****
VI	2244	0.23	0.23	*****	*****
II	2125	0.15	0.17	****	****
III	1901	0.20	0.21	*****	*****
VIII	1053	0.13	0.17	****	****
VII	310	0.35	0.38	*****	*****

2013	# observations	Investments Median	Investments Mean	all	core
I	3424	0.13	0.19	****	****
VI	2218	0.23	0.23	****	****
II	2135	0.16	0.19	****	****
III	1849	0.20	0.21	****	****
VIII	1424	0.11	0.15	****	****

2014	# observations	Investments Median	Investments Mean	all	core
I	3653	0.11	0.15	****	****
II	2390	0.18	0.21	***	****
VI	2086	0.23	0.23	****	****
III	1807	0.20	0.21	***	****
VII	226	0.45	0.47	****	****

2005	# observations	Customer Deposits Median	Mean	all	core
I	3221	0.44	0.42	***	***
II	2592	0.66	0.48	****	***
III	1946	0.82	0.79	****	****
IV	1416	0.57	0.57	****	****
V	606	0.58	0.43	***	****

2006	# observations	Customer Deposits Median	Mean	all	core
I	3390	0.33	0.37	****	****
II	2568	0.65	0.48	***	****
III	1952	0.82	0.79	****	****
IV	1534	0.58	0.58	***	***
V	463	0.63	0.54	**	***

2007	# observations	Customer Deposits Median	Mean	all	core
I	4108	0.41	0.40	**	**
II	3861	0.60	0.49	**	**
III	2070	0.81	0.78	**	**

2008	# observations	Customer Deposits Median	Mean	all	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	# observations	Customer Deposits Median	Mean	all	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	***	***

2010	# observations	Customer Deposits Median	Mean	all	core
I	3484	0.36	0.38	***	***
II	2760	0.41	0.36	***	***
VI	2233	0.75	0.72	***	***
III	2009	0.82	0.79	***	***

2011	# observations	Customer Deposits Median	Mean	all	core
I	3608	0.36	0.39	****	****
II	2692	0.38	0.34	****	****
VI	2252	0.75	0.72	****	****
III	1778	0.82	0.79	****	****
VIII	573	0.71	0.61	****	****

2012	# observations	Customer Deposits Median	Mean	all	core
I	3626	0.45	0.41	*****	*****
VI	2244	0.75	0.72	****	****
II	2125	0.14	0.24	*****	*****
III	1901	0.83	0.80	*****	*****
VIII	1053	0.76	0.72	****	****
VII	310	0.00	0.06	*****	*****

2013	# observations	Customer Deposits Median	Mean	all	core
I	3424	0.36	0.38	****	****
VI	2218	0.76	0.73	****	****
II	2135	0.23	0.27	****	****
III	1849	0.83	0.80	****	****
VIII	1424	0.71	0.64	****	****

2014	# observations	Customer Deposits Median	Mean	all	core
I	3653	0.53	0.46	****	****
II	2390	0.40	0.37	****	****
VI	2086	0.76	0.74	****	****
III	1807	0.83	0.80	****	****
VII	226	0.00	0.06	****	****

Interbank Borrowing

2005	*# observations*	*Median*	*Mean*	*all*	*core*
I	3221	0.16	0.24	****	****
II	2592	0.12	0.17	****	****
III	1946	0.02	0.06	****	****
IV	1416	0.03	0.08	***	***
V	606	0.05	0.09	***	***

Interbank Borrowing

2006	*# observations*	*Median*	*Mean*	*all*	*core*
I	3390	0.19	0.26	****	****
II	2568	0.12	0.17	***	****
III	1952	0.02	0.06	****	****
IV	1534	0.03	0.07	****	****
V	463	0.08	0.11	***	****

Interbank Borrowing

2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.15	0.24	**	**
II	3861	0.09	0.14	**	**
III	2070	0.03	0.06	**	**

Interbank Borrowing

2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.19	0.25	***	***
II	3003	0.04	0.14	***	**
VI	2053	0.11	0.13	***	***
III	1708	0.02	0.05	***	**

Interbank Borrowing

2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.17	0.24	***	***
II	2877	0.02	0.14	***	***
VI	2195	0.14	0.15	***	***
III	2125	0.03	0.06	***	***

Interbank Borrowing

2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.15	0.23	***	***
II	2760	0.04	0.15	***	**
VI	2233	0.13	0.14	***	***
III	2009	0.02	0.06	***	**

Interbank Borrowing

2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.14	0.21	****	****
II	2692	0.11	0.19	***	****
VI	2252	0.12	0.13	****	****
III	1778	0.02	0.06	****	****
VIII	573	0.06	0.17	***	****

Interbank Borrowing

2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.14	0.21	*****	*****
VI	2244	0.12	0.13	****	*****
II	2125	0.17	0.22	****	*****
III	1901	0.02	0.05	*****	*****
VIII	1053	0.04	0.08	*****	*****
VII	310	0.36	0.38	*****	*****

Interbank Borrowing

2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.15	0.22	****	****
VI	2218	0.10	0.12	***	****
II	2135	0.16	0.21	***	****
III	1849	0.02	0.05	****	****
VIII	1424	0.05	0.11	****	****

Interbank Borrowing

2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.11	0.17	***	***
II	2390	0.11	0.18	***	****
VI	2086	0.10	0.12	**	***
III	1807	0.02	0.05	****	****
VII	226	0.43	0.44	****	****

		Long-Term Funding			
2005	*# observations*	*Median*	*Mean*	*all*	*core*
I	3221	0.03	0.09	***	***
II	2592	0.01	0.04	***	****
III	1946	0.00	0.03	***	****
IV	1416	0.14	0.18	****	****
V	606	0.06	0.10	***	***

		Long-Term Funding			
2006	*# observations*	*Median*	*Mean*	*all*	*core*
I	3390	0.04	0.10	****	****
II	2568	0.01	0.04	****	****
III	1952	0.00	0.03	****	****
IV	1534	0.11	0.17	***	****
V	463	0.11	0.14	***	****

		Long-Term Funding			
2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.04	0.11	**	**
II	3861	0.03	0.11	**	**
III	2070	0.00	0.03	**	**

		Long-Term Funding			
2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.04	0.11	**	**
II	3003	0.03	0.13	*	***
VI	2053	0.03	0.07	**	**
III	1708	0.00	0.03	***	***

		Long-Term Funding			
2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.05	0.12	***	***
II	2877	0.01	0.12	**	***
VI	2195	0.02	0.05	**	***
III	2125	0.00	0.03	***	***

		Long-Term Funding			
2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.04	0.11	***	***
II	2760	0.01	0.13	***	**
VI	2233	0.01	0.04	***	**
III	2009	0.00	0.03	***	***

		Long-Term Funding			
2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.04	0.12	****	****
II	2692	0.02	0.11	****	***
VI	2252	0.01	0.04	****	***
III	1778	0.00	0.02	****	****
VIII	573	0.00	0.02	****	****

		Long-Term Funding			
2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.04	0.12	*****	*****
VI	2244	0.01	0.04	****	****
II	2125	0.02	0.12	****	****
III	1901	0.00	0.02	*****	*****
VIII	1053	0.00	0.01	*****	*****
VII	310	0.01	0.05	***	***

		Long-Term Funding			
2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.03	0.12	****	****
VI	2218	0.01	0.04	***	****
II	2135	0.02	0.11	****	****
III	1849	0.00	0.02	****	***
VIII	1424	0.00	0.06	***	***

		Long-Term Funding			
2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.04	0.12	****	****
II	2390	0.01	0.09	***	**
VI	2086	0.00	0.04	***	***
III	1807	0.00	0.02	****	****
VII	226	0.02	0.05	**	***

2005	# observations	LT Fund. and Equity Median	Mean	all	core	2006	# observations	LT Fund. and Equity 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	****	****

2007	# observations	LT Fund. and Equity Median	Mean	all	core	2008	# observations	LT Fund. and Equity 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	**	**

2009	# observations	LT Fund. and Equity Median	Mean	all	core	2010	# observations	LT Fund. and Equity 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	***	***

2011	# observations	LT Fund. and Equity Median	Mean	all	core	2012	# observations	LT Fund. and Equity 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	***	****

2013	# observations	LT Fund. and Equity Median	Mean	all	core	2014	# observations	LT Fund. and Equity 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			
2005	*# observations*	*Median*	*Mean*	*all*	*core*
I	3221	0.63	0.57	****	***
II	2592	0.72	0.53	****	***
III	1946	0.87	0.84	****	****
IV	1416	0.83	0.78	****	****
V	606	0.77	0.55	****	**

		Stable Funding			
2006	*# observations*	*Median*	*Mean*	*all*	*core*
I	3390	0.58	0.53	***	****
II	2568	0.71	0.53	***	****
III	1952	0.87	0.84	****	****
IV	1534	0.83	0.79	****	***
V	463	0.82	0.70	****	***

		Stable Funding			
2007	*# observations*	*Median*	*Mean*	*all*	*core*
I	4108	0.64	0.61	**	**
II	3861	0.76	0.62	**	**
III	2070	0.87	0.84	**	**

		Stable Funding			
2008	*# observations*	*Median*	*Mean*	*all*	*core*
I	3270	0.57	0.53	***	***
II	3003	0.71	0.56	***	***
VI	2053	0.79	0.74	***	***
III	1708	0.88	0.85	***	***

		Stable Funding			
2009	*# observations*	*Median*	*Mean*	*all*	*core*
I	3116	0.57	0.52	***	***
II	2877	0.66	0.52	***	***
VI	2195	0.77	0.76	***	***
III	2125	0.86	0.83	***	***

		Stable Funding			
2010	*# observations*	*Median*	*Mean*	*all*	*core*
I	3484	0.61	0.54	**	***
II	2760	0.66	0.52	**	***
VI	2233	0.78	0.76	***	***
III	2009	0.86	0.83	***	***

		Stable Funding			
2011	*# observations*	*Median*	*Mean*	*all*	*core*
I	3608	0.63	0.56	****	****
II	2692	0.58	0.48	****	****
VI	2252	0.78	0.76	****	****
III	1778	0.86	0.83	****	****
VIII	573	0.76	0.66	****	****

		Stable Funding			
2012	*# observations*	*Median*	*Mean*	*all*	*core*
I	3626	0.65	0.57	*****	*****
VI	2244	0.79	0.77	****	*****
II	2125	0.46	0.39	****	****
III	1901	0.86	0.84	*****	*****
VIII	1053	0.80	0.77	****	*****
VII	310	0.25	0.32	****	****

		Stable Funding			
2013	*# observations*	*Median*	*Mean*	*all*	*core*
I	3424	0.63	0.56	****	****
VI	2218	0.79	0.78	****	***
II	2135	0.49	0.41	****	****
III	1849	0.86	0.84	****	****
VIII	1424	0.79	0.73	****	***

		Stable Funding			
2014	*# observations*	*Median*	*Mean*	*all*	*core*
I	3653	0.69	0.62	****	****
II	2390	0.60	0.48	****	****
VI	2086	0.79	0.78	****	****
III	1807	0.86	0.83	****	****
VII	226	0.15	0.20	****	****

2005	# observations	Stable Funding (core) Median	Mean	all	core
I	3221	0.55	0.50	***	***
II	2592	0.67	0.49	***	***
III	1946	0.84	0.81	****	****
IV	1416	0.80	0.75	****	****
V	606	0.73	0.53	****	**

2006	# observations	Stable Funding (core) Median	Mean	all	core
I	3390	0.50	0.46	****	****
II	2568	0.66	0.50	****	****
III	1952	0.84	0.81	****	****
IV	1534	0.81	0.75	****	***
V	463	0.79	0.67	****	***

2007	# observations	Stable Funding (core) Median	Mean	all	core
I	4108	0.55	0.50	**	**
II	3861	0.71	0.58	**	**
III	2070	0.84	0.81	**	**

2008	# observations	Stable Funding (core) Median	Mean	all	core
I	3270	0.48	0.46	***	***
II	3003	0.66	0.53	***	***
VI	2053	0.75	0.70	***	***
III	1708	0.85	0.82	***	***

2009	# observations	Stable Funding (core) Median	Mean	all	core
I	3116	0.49	0.46	***	***
II	2877	0.60	0.49	***	***
VI	2195	0.75	0.73	***	***
III	2125	0.84	0.81	***	***

2010	# observations	Stable Funding (core) Median	Mean	all	core
I	3484	0.54	0.48	***	***
II	2760	0.60	0.48	***	***
VI	2233	0.76	0.74	***	***
III	2009	0.84	0.81	***	***

2011	# observations	Stable Funding (core) Median	Mean	all	core
I	3608	0.56	0.50	****	****
II	2692	0.53	0.44	****	****
VI	2252	0.76	0.74	****	****
III	1778	0.84	0.81	****	****
VIII	573	0.72	0.63	****	****

2012	# observations	Stable Funding (core) Median	Mean	all	core
I	3626	0.59	0.52	*****	*****
VI	2244	0.77	0.75	****	****
II	2125	0.38	0.35	*****	****
III	1901	0.84	0.82	*****	*****
VIII	1053	0.77	0.73	****	****
VII	310	0.05	0.11	*****	****

2013	# observations	Stable Funding (core) Median	Mean	all	core
I	3424	0.57	0.49	****	****
VI	2218	0.78	0.76	****	****
II	2135	0.42	0.37	****	****
III	1849	0.84	0.82	****	****
VIII	1424	0.76	0.68	****	****

2014	# observations	Stable Funding (core) Median	Mean	all	core
I	3653	0.63	0.57	****	****
II	2390	0.55	0.45	****	****
VI	2086	0.78	0.77	****	****
III	1807	0.84	0.81	****	****
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	****	****

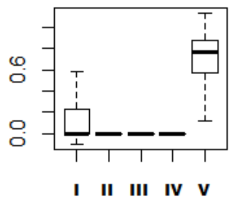
C.3.2 Multiple Comparisons: examples

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.

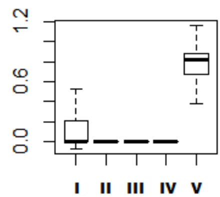
In the first example, we consider *Retail Loans* in 2005 and 2006. Distributions are quite similar in both years. Group *V* has very different median/mean values. We remove this group and we test multiple pairwise comparisons among the remaining four groups. Results confirm the same statistically significant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

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	****	****		



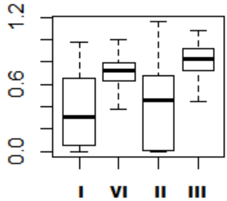
2006	*# obs.*	*Median*	*Mean*	*all*	*core*	I-VI	I-VI (core)
I	3390	0.00	0.14	****	****	***	***
II	2568	0.00	0.01	****	***	***	**
III	1952	0.00	0.02	****	****	***	***
IV	1534	0.00	0.01	****	***	***	**
V	463	0.81	0.76	****	****		



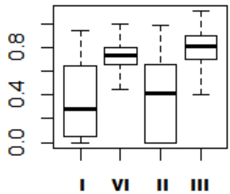
In the second example, we focus on *Customer Deposits* in 2008 and 2009. Even in this case distributions are quite similar in both years, although in 2009 groups *I* and *II* have not statistically different medians (in the *I-II* specification). However, we observe that group *I* and *II* are more similar 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 C.3.1.

Customer Dep.

2008	# obs.	Median	Mean	all	core	I-II	I-II (core)	III-VI
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
I	3116	0.29	0.35	**	***		*	
II	2877	0.41	0.37	**	***		*	
VI	2195	0.74	0.71	***	***			*
III	2125	0.81	0.78	***	***			*



C.3.3 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 4.3.1 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 4.4 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 43: 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**
I										
Max.	0.45	0.50	0.73	0.47	0.56	0.50	0.55	0.46	0.55	0.62
Mean	0.08	0.09	0.06	0.09	0.09	0.09	0.08	0.08	0.08	0.07
Median	0.07	0.07	0.06	0.07	0.07	0.07	0.07	0.06	0.06	0.06
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
II										
Max.	0.67	0.64	0.50	0.50	0.60	0.45	0.55	0.42	0.45	0.55
Mean	0.07	0.07	0.06	0.07	0.08	0.07	0.08	0.07	0.08	0.08
Median	0.00	0.00	0.05	0.06	0.06	0.06	0.07	0.06	0.06	0.06
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
III										
Max.	0.67	0.55	0.50	0.55	0.64	0.58	0.50	0.46	0.47	0.50
Mean	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Median	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IV										
Max.	0.50	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.07	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V										
Max.	0.69	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.07	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VI										
Max.	0.00	0.00	0.00	0.73	0.75	0.67	0.64	0.55	0.56	0.54
Mean	0.00	0.00	0.00	0.08	0.08	0.08	0.09	0.08	0.09	0.09
Median	0.00	0.00	0.00	0.06	0.05	0.05	0.05	0.05	0.05	0.06
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.36
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.10
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.08
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VIII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.55	0.42	0.00
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.08	0.08	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.07	0.00
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 44: 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**
I										
Max.	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.66
Mean	0.45	0.45	0.44	0.44	0.42	0.42	0.40	0.38	0.38	0.40
Median	0.48	0.48	0.48	0.48	0.48	0.48	0.45	0.41	0.41	0.41
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
II										
Max.	0.66	0.66	0.69	0.69	0.66	0.66	0.66	0.69	0.66	0.66
Mean	0.50	0.50	0.49	0.49	0.50	0.49	0.48	0.47	0.47	0.48
Median	0.45	0.45	0.48	0.48	0.52	0.52	0.52	0.52	0.52	0.52
Min.	0.00	0.14	0.03	0.03	0.00	0.07	0.00	0.03	0.10	0.07
III										
Max.	0.62	0.66	0.66	0.66	0.66	0.62	0.62	0.66	0.62	0.62
Mean	0.43	0.42	0.42	0.40	0.38	0.37	0.36	0.34	0.34	0.35
Median	0.38	0.38	0.38	0.38	0.34	0.34	0.31	0.31	0.31	0.31
Min.	0.07	0.07	0.00	0.00	0.07	0.00	0.00	0.03	0.07	0.03
IV										
Max.	0.66	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.50	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.52	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.10	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V										
Max.	0.62	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.46	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.45	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VI										
Max.	0.00	0.00	0.00	0.69	0.62	0.66	0.62	0.66	0.66	0.62
Mean	0.00	0.00	0.00	0.37	0.34	0.34	0.34	0.33	0.34	0.40
Median	0.00	0.00	0.00	0.38	0.28	0.31	0.31	0.31	0.34	0.41
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07
VII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.00	0.69
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.46
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.00	0.48
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
VIII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.66	0.66	0.00
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.48	0.46	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.48	0.45	0.00
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 45: 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**
I										
Max.	0.79	0.83	0.90	0.79	0.86	0.83	0.83	0.79	0.83	0.83
Mean	0.49	0.49	0.47	0.48	0.47	0.47	0.45	0.42	0.43	0.43
Median	0.48	0.52	0.48	0.52	0.48	0.52	0.48	0.45	0.45	0.45
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
II										
Max.	0.86	0.86	0.79	0.83	0.86	0.79	0.83	0.76	0.79	0.83
Mean	0.53	0.53	0.52	0.52	0.54	0.53	0.52	0.51	0.51	0.52
Median	0.48	0.48	0.48	0.52	0.55	0.55	0.55	0.55	0.55	0.55
Min.	0.21	0.21	0.03	0.07	0.07	0.14	0.14	0.10	0.14	0.07
III										
Max.	0.86	0.83	0.79	0.83	0.86	0.83	0.79	0.76	0.76	0.76
Mean	0.47	0.47	0.46	0.45	0.43	0.42	0.41	0.39	0.39	0.40
Median	0.45	0.45	0.45	0.41	0.38	0.38	0.34	0.34	0.34	0.34
Min.	0.17	0.21	0.17	0.14	0.07	0.07	0.07	0.07	0.10	0.03
IV										
Max.	0.79	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.54	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.55	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.10	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V										
Max.	0.83	0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	0.50	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.48	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min.	0.17	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VI										
Max.	0.00	0.00	0.00	0.90	0.90	0.86	0.86	0.83	0.83	0.79
Mean	0.00	0.00	0.00	0.42	0.39	0.39	0.39	0.38	0.39	0.45
Median	0.00	0.00	0.00	0.41	0.34	0.34	0.34	0.34	0.38	0.45
Min.	0.00	0.00	0.00	0.14	0.17	0.14	0.17	0.17	0.14	0.14
VII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.00	0.79
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.51	0.00	0.51
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.55
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
VIII										
Max.	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.83	0.76	0.00
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.52	0.49	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.55	0.52	0.00
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00

C.3.4 Mixed Groups

We briefly comment here on the characteristics of three mixed groups adopted by a very volatile set of institutions worldwide (see Section 4.5.3 for details.) during the periods 2011-2013. The liability side shows a well mixed composition between wholesale and costumer deposits, in particular in 2011 and 2013, that places the funding strategy of those institutions in between a deposit-oriented and wholesale-oriented business models. The asset 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 retail and commercial loan investment strategy.

Table 46: Mixed groups. We report average values for aggregated balance sheet variables standardized by total assets for the Mixed groups. For variables definitions see Appendix C.2. Estimates are computed for each year over the interval 2011-13. Last row provides summary statistics for Total Assets (in USD Billion).

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

Bibliography

[1] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. "Cascades in networks and aggregate volatility". In: *National Bureau of Economic Research* (2010).

[2] V. Acharya, T. F. Cooley, M. P. Richardson, and I. Walter. *Regulating Wall Street: The Dodd-Frank Act and the new architecture of global finance*. Vol. 608. John Wiley & Sons, 2010.

[3] V. Acharya, L. Pedersen, T. Philippon, and M. Richardson. "Regulating systemic risk". In: *Restoring financial stability: How to repair a failed system* (2009), pp. 283–304.

[4] T. Adrian and H.S. Shin. "Liquidity and leverage". In: *Journal of Financial Intermediation* 19 (2008), pp. 418–437.

[5] F. Allen and A. Babus. "Networks in finance". In: *Wharton Financial Institutions Center Working Paper* (2008).

[6] F. Allen and D. Gale. "Financial contagion". In: *Journal of political economy* 108.1 (2000), pp. 1–33.

[7] F. Allen, X. Gu, and O. Kowalewski. "Financial crisis, structure and reform". In: *Journal of Banking & Finance* 36 (2012), pp. 2960–2973.

[8] Y. Altunbas, S. Manganelli, and D. Marques-Ibanez. "Bank Risk During the Financial Crisis. Do Business Model Matter?" In: *European Central Bank* (2011).

[9] B. E. Aronson. "A Reassessment of Japan's Big Bang Financial Regulatory Reform". In: *IMES Discussion Paper Series* 2011-E-19 (2011).

[10] M. Avellaneda and R. Cont. "Transparency in Over-the-counter Interest rate derivatives Markets". In: *ISDA, Finance Concepts* (2010).

[11] R. Ayadi, E. Arbak, and P. W. De Groen. “Business models in European banking: A pre-and post-crisis screening”. In: *Center for European Policy Studies* (2011).

[12] R. Ayadi, E. Arbak, and P. W. De Groen. “Regulation of European banks and business models: towards a new paradigm?” In: *Centre for European Policy Studies* (2012).

[13] R. Ayadi and P. W. De Groen. “Banking business models 2015 Europe”. In: *Alphonse and Dorimène Desjardins International Institute for Cooperatives and International Research Centre on Cooperative Finance* (2015).

[14] J. A. Batten and P. G. Szilagyi. “Why Japan Needs to Develop its Corporate Bond Market”. In: *International Journal of the Economics of Business* 10(1) (2003),

[15] S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz. “Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk”. In: *Journal of Economic Dynamics and Control* 36.8 (2012), pp. 1121–1141.

[16] BCBS. “An Assessment of the Long-Term Economic Impact of Stronger Capital and Liquidity Requirements”. In: *Bank of International Settlements* (2010).

[17] BCBS. “Basel 3: A global regulatory framework for more resilient banks and banking systems”. In: *Bank of International Settlements* (2011).

[18] BCBS. “Capital requirements for banks’ equity investments in funds”. In: *Bank of International Settlements* (2013).

[19] BCBS. “Global systemically important banks: Assessment methodology and the additional loss absorbency requirement”. In: *Bank of International Settlements* (2011).

[20] T. Beck. “The Econometrics of Finance and Growth”. In: *In Palgrave Handbook of Econometrics* 2 (2009), pp. 1180–1211.

[21] T. Beck. “The Role of Finance in Economic Development: Benefits, Risks, and Politics”. In: *European Banking Center Discussion Paper* 2011-038 (2011).

[22] A. Beltratti and R. M. Stultz. “The credit crisis around the globe: Why did some banks perform better?” In: *Journal of Financial Economics* 105.I1 (2012), pp. 1–17.

[23] G. J. Benston. “Universal Banking”. In: *Journal of Economic Perspectives* 8.3 (1994), pp. 121–144.

[24] S. Bentolila, P. Cahuc, J. J. Dolado, and T. Le Barbanchon. “Two-tier labour markets in the Great Recession: France versus Spain”. In: *The Economic Journal* 122.562 (2012), F155–F187.

[25] A. N. Berger and C. H. S. Bouwman. “How does capital affect bank performance during financial crises?” In: *Journal of Financial Economics* 109 (2013), pp. 146–176.

[26] F. Betz, S. Oprică, T. A. Peltonen, and P. Sarlin. “Predicting distress in European banks”. In: *Journal of Banking & Finance* 45 (2014), pp. 225–241.

[27] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. “Econometric measures of connectedness and systemic risk in the finance and insurance sectors”. In: *Journal of Financial Economics* 104.3 (2012), pp. 535–559.

[28] BIS. “Bank of International Settlements 84th Annual Report (June 2014)”. In: *Bank of International Settlements* (2014).

[29] BIS. “Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement”. In: *Basel Committee on Banking Supervision* 255 (2013).

[30] D. Bisias, M.D. Flood, A.W. Lo, and S. Valavanis. “A survey of systemic risk analytics”. In: *US Department of Treasury, Office of Financial Research* 0001 (2012).

[31] V. D. Blondel, J-L. Guillaume, R. Lambiotte, and E. Lefebvre. “Fast unfolding of communities in large networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (2008), P10008+.

[32] A. Blundell-Wignall and C. Roulet. “Business models of banks, leverage and the distance-to-default”. In: *OECD Journal: Financial Market Trends* Vol. 2012/2 (2013).

[33] P. Bongini, L. Nieri, and M. Pelagatti. “The importance of being systemically important financial institutions”. In: *Journal of Banking & Finance* 50 (2015), pp. 562–574.

[34] M. Bonollo, I. Crimaldi, A. Flori, F. Pammolli, and M. Riccaboni. “Systemic risk and banking regulation: some facts on the new regulatory framework”. In: *Corporate Ownership & Control* 12(2) (2015).

[35] M. Boss, H. Elsinger, M. Summer, and S Thurner. "Network topology of the interbank market". In: *Quantitative Finance* 4.6 (2004), pp. 677–684.

[36] M.K. Brunnermeier. "Deciphering the liquidity and credit crunch 2007-2008". In: *Journal of Economic Perspectives* 23 (2009), pp. 77–100.

[37] M.K. Brunnermeier and M. Oehmke. "Bubbles, financial crises, and systemic risk". In: *National Bureau of Economic Research* (2012).

[38] T. Caliński and J. Harabasz. "A dendrite method for cluster analysis". In: *Communications in Statistics-theory and Methods* 3.1 (1974), pp. 1–27.

[39] C. W. Calomiris and D. Nissim. "Crisis-related shifts in the market valuation of banking activities". In: *Journal of Financial Intermediation* 23 (2014), pp. 400–435.

[40] S. G. Cecchetti, J. Gyntelberg, and M. Hollanders. "Central counterparties for over-the-counter derivatives". In: *BIS Quarterly Review, September* (2009).

[41] T. Chakraborty, S. Srinivasan, N. Ganguly, S. Bhowmick, and A. Mukherjee. "Constant communities in complex networks". In: *Scientific reports* 3 (2013).

[42] S.A. Corwin and P. Schultz. "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices". In: *The Journal of Finance* 67(2) (2012).

[43] O. De Bandt, P. Hartmann, and J. L. Peydró. "Systemic risk in banking: an update". In: *The Oxford Handbook of Banking*. Oxford University Press, 2009.

[44] H. Degryse and G. Nguyen. "Interbank exposures: An empirical examination of contagion risk in the Belgian banking system". In: *International Journal of Central Banking* 3.2 (2007), pp. 123–171.

[45] G. Dell'Ariccia, E. Detragiache, and R. Rajan. "The real effect of banking crises". In: *Journal of Financial Intermediation* 17 (2008), pp. 89–112.

[46] A. Demirgüç-Kunt and E. Detragiache. "Cross-country empirical studies of systemic bank distress: a survey". In: *National Institute Economic Review* 192.1 (2005), pp. 68–83.

[47] A. Demirgüc-Kunt and H.P. Huizinga. “Bank Activity and Funding Strategies: The Impact on Risk and Return”. In: *Journal of Financial Economics* 98.Issue 3 (2010), pp. 626–650.

[48] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the royal statistical society. Series B (methodological)* (1977), pp. 1–38.

[49] Y. Demyanyk and I. Hasan. “Financial crises and bank failures: A review of prediction methods”. In: *Omega* 38.5 (2010), pp. 315–324.

[50] P. Deuskar, A. Gupta, and M. G. Subrahmanyam. “Liquidity effect in OTC options markets: Premium or discount?” In: *Journal of Financial Markets* 14(1) (2011).

[51] D.W. Diamond and R.G. Rajan. “The credit crisis: conjectures about causes and remedies”. In: *American Economic Review* 99 (2009), pp. 606–610.

[52] S. van Dongen and A. J. Enright. “Metric distances derived from cosine similarity and Pearson and Spearman correlations”. In: *CoRR* abs/1208.3145 (2012).

[53] Depository Trust & Clearing Corporation DTCC. “Beyond the Horizon, A White Paper to the Industry on Systemic Risk”. In: *DTCC white paper* (2013).

[54] Depository Trust & Clearing Corporation DTCC. “DTCC Data Repository (U.S.) LLC Rulebook”. In: *Rulebook revised as of 10/03/2014* (2014).

[55] D. Duffie, A. Li, and T. Lubke. “Policy perspectives on OTC derivatives market infrastructure”. In: *FRB of New York Staff Report* 424 (2010).

[56] D. Duffie and H. Zhu. “Does a central clearing counterparty reduce counterparty risk?” In: *Review of Asset Pricing Studies* 1.1 (2011), pp. 74–95.

[57] T. Duprey and M. Lé. “Bankscope dataset: getting started”. In: *Available at SSRN 2191449* (2014).

[58] EBA. “Guidelines: on common procedures and methodologies for the supervisory review and evaluation process (SREP)”. In: *European Banking Authority* (2014).

[59] European Commission EC. “Delegated Regulation on the data to be published and made available by trade repositories and operational standards for aggregating, comparing and accessing the data”. In: *Delegated Regulation 151* (2013).

[60] European Commission EC. “Delegated Regulation on the minimum details of the data to be reported to trade repositories”. In: *Delegated Regulation 148* (2013).

[61] ECB. “Banking Structures Report (November 2013)”. In: *European Central Bank* (2013).

[62] L. Eisenberg and T. H. Noe. “Systemic risk in financial systems”. In: *Management Science* 47.2 (2001), pp. 236–249.

[63] M. Elliott, B. Golub, and M. O. Jackson. “Financial networks and contagion”. In: *The American economic review* 104.10 (2014), pp. 3115–3153.

[64] H. Elsinger, A. Lehar, and M. Summer. “Risk assessment for banking systems”. In: *Management science* 52.9 (2006), pp. 1301–1314.

[65] European Union Parliament EUP. “European Market Infrastructure Regulation”. In: *Regulation 648/2012* (2012).

[66] FASB. “Comparability in International Accounting Standards: A Brief History”. In: *Financial Accounting Standards Board* . Retrieved on March 8 (2016).

[67] Financial Stability Board. “Feasibility Study on Approaches to Aggregate OTC Derivatives Data”. In: *Consultation Paper (Basel: Bank for International Settlements)* (2014).

[68] Financial Stability Board. “OTC Derivatives Market Reforms”. In: *FSB, Ninth Progress Report on Implementation* (2015).

[69] Financial Stability Board. “Reducing the moral hazard posed by systemically important financial institutions”. In: *FSB Recommendations and Time Lines* (2010).

[70] D. Firth. “Bias reduction of maximum likelihood estimates”. In: *Biometrika* 80.1 (1993), pp. 27–38.

[71] M. J. Flannery. “Using market information in prudential bank supervision: A review of the US empirical evidence”. In: *Journal of Money, Credit and Banking* (1998), pp. 273–305.

[72] M. J. Fleming, J. P. Jackson, A. Li, A. Sarkar, and P. Zobel. "An analysis of OTC interest rate derivatives transactions: implications for public reporting". In: *FRB of New York Staff Report* 557 (2012).

[73] S. Fortunato. "Community detection in graphs". In: *Physics Reports* 486.3-5 (2010), pp. 75–174.

[74] X. Freixas, B. M. Parigi, and J-C. Rochet. "Systemic risk, interbank relations, and liquidity provision by the central bank". In: *Journal of money, credit and banking* (2000), pp. 611–638.

[75] C. Furfine. "Interbank exposures: Quantifying the risk of contagion". In: *Journal of Money, Credit, and Banking* 35.1 (2003), pp. 111–128.

[76] P. Gai and S. Kapadia. "Contagion in financial networks". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. The Royal Society. 2010.

[77] M. Gehde-Trapp, Y. Gunduz, and J. Nasev. "The liquidity premium in CDS transaction prices: Do frictions matter?" In: *Journal of Banking & Finance* (2015).

[78] B. González-Hermosillo. "Determinants of ex-ante banking system distress: A macro-micro empirical exploration of some recent episodes". In: *International Monetary Fund* (1999).

[79] E. Gurvich and I. Prilepskiy. "The impact of financial sanctions on the Russian economy". In: *Russian Journal of Economics* 1.4 (2015), pp. 359–385.

[80] A. G. Haldane. "Rethinking the financial network". In: *Fragile stabilität–stabile fragilität*. Springer, 2013, pp. 243–278.

[81] A. G. Haldane and R. M. May. "Systemic risk in banking ecosystems". In: *Nature* 469.7330 (2011), pp. 351–355.

[82] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. "On clustering validation techniques". In: *Journal of intelligent information systems* 17.2 (2001), pp. 107–145.

[83] J. Han, M. Kamber, and J. Pei. *Data mining: concepts and techniques*. Elsevier, 2011.

[84] N. Hautsch, J. Schaumburg, and M. Schienle. "Financial network systemic risk contributions". In: *Review of Finance* (2014).

[85] W. Hoeffding. "Scale-Invariant Correlation Theory". In: *The collected works of Wassily Hoeffding*. Springer, 1994, pp. 57–107.

[86] J. Honaker and G. King. “What to do about missing values in time-series cross-section data”. In: *American Journal of Political Science* 54.2 (2010), pp. 561–581.

[87] T. Hoshi and A. Kashyap. “The Japanese Banking Crisis: Where did it come from and how will it end?” In: *NBER Working Paper Series* (1999).

[88] H.P. Huizinga and L. Laeven. “Bank valuation and accounting discretion during a financial crisis”. In: *Journal of Financial Economics* 106 (2012), pp. 614–634.

[89] J. Hull. “The changing landscape for derivatives”. In: *Journal of Financial Engineering* 1.03 (2014), p. 1450021.

[90] V. Huynh, V. Kreinovich, and S. Sriboonchitta. *Modeling Dependence in Econometrics*. Springer, 2014.

[91] IMF-BIS-FSB. “Report to G-20 Finance Ministers and Governors. Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations”. In: *International Monetary Fund and Bank for International Settlement and Financial Stability Board* (2009).

[92] G. Iori, G. De Masi, O. V. Precup, G. Gabbi, and G. Caldarelli. “A network analysis of the Italian overnight money market”. In: *Journal of Economic Dynamics and Control* 32.1 (2008), pp. 259–278.

[93] I. Jeffries. *Economic developments in contemporary Russia*. Taylor & Francis, 2011.

[94] H. Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[95] T. Kick and M. Koetter. “Slippery slopes of stress: ordered failure events in German banking”. In: *Journal of Financial Stability* 3.2 (2007), pp. 132–148.

[96] M. Köhler. “Which banks are more risky? The impact of business models on bank stability”. In: *Journal of Financial Stability* 16 (2015), pp. 195–212.

[97] L. Laeven and F. Valencia. *Resolution of banking crises: The good, the bad, and the ugly*. 10-146. International Monetary Fund, 2010.

[98] L. Laeven and F. Valencia. “Systemic banking crises database: An update”. In: (2012).

[99] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato. “Finding statistically significant communities in networks”. In: *PloS one* 6.4 (2011), e18961.

[100] S. Langfield, Z. Liu, and T. Ota. “Mapping the UK interbank system”. In: *Journal of Banking & Finance* 45 (2014), pp. 288–303.

[101] K. Lannoo. “The New Financial Regulatory Paradigm: A transatlantic perspective”. In: *CEPS Policy Briefs* 287 (2013).

[102] R. Levine. “Finance and Growth: Theory and Evidence”. In: *In Handbook of Economic Growth* (2005), pp. 865–934.

[103] A. Lozano-Vivas and F. Pasiouras. “The impact of non-traditional activities on the estimation of bank efficiency: international evidence”. In: *Journal of Banking & Finance* 34.7 (2010), pp. 1436–1449.

[104] M. Mahfoud. “Bivariate Archimedean copulas: an application to two stock market indices”. In: *BMI Paper* (2012).

[105] S. Markose, S. Giansante, and A. R. Shaghaghi. “Too interconnected to fail: financial network of US CDS market: Topological fragility and systemic risk”. In: *Journal of Economic Behavior & Organization* 83.3 (2012), pp. 627–646.

[106] F. Mergaerts and R. Vander Vennet. “Business models and bank performance: A long-term perspective”. In: *Journal of Financial Stability* 22 (2016), pp. 57–75.

[107] P. E. Mistrulli. “Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns”. In: *Journal of Banking & Finance* 35.5 (2011), pp. 1114–1127.

[108] R. B. Nelsen. “An introduction to copulas”. In: *Lecture Notes in Statistics. New York: Springer* (2006).

[109] M. E. J. Newman and M. Girvan. “Finding and evaluating community structure in networks”. In: *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* 69.2 Pt 2 (2004), p. 16.

[110] A. Pallavicini and D. Brigo. “Interest-rate modelling in collateralized markets: multiple curves, credit-liquidity effects”. In: *CCPs. arxiv. org, ssrn. com* (2013).

[111] Z. Pozsar, T. Adrian, A. Ashcraft, and H. Boesky. “Shadow Banking”. In: *Federal Reserve Bank of New York Staff Reports* (2012).

[112] C. M. Reinhart and K. S. Rogoff. “This Time Is Different: Eight Centuries of Financial Folly”. In: *Princeton, N.J.: Princeton University Press* (2009).

[113] R. Roengpitya, N. Tarashev, and K. Tsatsaronis. “Bank business models”. In: *BIS Quarterly Review* (2014).

[114] D. Russo. “OTC derivatives: financial stability challenges and responses from authorities”. In: *FSR FINANCIAL* (2010), p. 101.

[115] M. Segoviano and C. Goodhart. “Bank stability measures”. In: *IMF, WP/09/4* (2009).

[116] H.S. Shin. “Global banking glut and loan risk premium”. In: *IMF Economic Review* 60.2 (2012), pp. 155–192.

[117] M. Sklar. *Fonctions de répartition a n dimensions et leurs marges*. Université Paris 8, 1959.

[118] J. Slive, J. Witmer, and E. Woodman. “Liquidity and central clearing: Evidence from the CDS market”. In: *Bank of Canada Working Paper* (2012).

[119] P-N. Tan, M. Steinbach, and V. Kumar. *Introduction to data mining*. Vol. 1. Pearson Addison Wesley Boston, 2006.

[120] V. A. Traag, G. Krings, and P. Van Dooren. “Significant Scales in Community Structure”. In: *Scientific Reports* 3 (2013),

[121] P. Trivedi and D. Zimmer. *Copula modeling: an introduction for practitioners*. Now Publishers Inc., 2007.

[122] C. Upper. “Simulation methods to assess the danger of contagion in interbank markets”. In: *Journal of Financial Stability* 7.3 (2011), pp. 111–125.

[123] US 111th Congress. “Dodd-Frank Wall Street Reform and Consumer Protection Act”. In: *Public Law 111-203* (2010).

[124] D. Valiante. “Shaping Reforms and Business Models for the OTC Derivatives Market: Quo vadis?” In: *ECMI Research Report* (2010).

[125] F. Vazquez and P. Federico. “Bank funding structures and risk: Evidence from the global financial crisis”. In: *Journal of Banking & Finance* 61 (2015), pp. 1–14.

[126] VIP. *Product Review of BvD Bankscope, October 2011*. 2011.

[127] J. Ward and H. Joe. “Hierarchical grouping to optimize an objective function”. In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244.