The Quest for Accurate Systemic Stress Tests

Amanah Ramadiah

First Supervisor: Dr. Fabio Caccioli

Second Supervisor: Dr. Daniel Fricke

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science

University College London

September 22, 2020

I, Amanah Ramadiah, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Any views expressed in this thesis are solely those of mine and so cannot be taken to represent those of Bank of England or to state Bank of England policy. This thesis should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee or Prudential Regulation Committee.

Abstract

After the global financial crisis of 2007-09, network approaches have become more prominent for analysing systemic risk in financial networks. Accordingly, various systemic stress testing models have been introduced in the literature. However, relatively little work has been devoted to study the accuracy of these models to predict systemic risk. This is especially important, given that stress tests are subjected to a range of modelling choices. In this thesis, I address this gap by studying the impact of modelling choices on the outcome of stress tests. I mainly focus on indirect contagion channel due to common assets holdings (overlapping portfolios) between financial institutions. In particular, this thesis is concerned with three aspects of stress tests related to indirect contagion: (i) the network of interactions, (ii) the dynamics of contagion, and (iii) the perimeter, or the types of institutions that are included in the stress test.

With regard to the first aspect, the network of financial linkages between financial institutions is often lacking, and one has to resort to network reconstruction methods to infer the network from partial information. In this thesis, I conduct a horse race between different network reconstruction methods in terms of their ability to reproduce the topological features and the levels of systemic risk of the actual bank-firm credit networks in Japan. I find that there is no single ”best” network reconstruction method - it depends on the assumed criterion of interest, but the reconstruction method which preserves the actual degree distribution overall consistently performs best. Furthermore, I find that the observed credit network displays relatively high levels of systemic risk compared with most reconstruction methods.

Concerning the second aspect, the propagation of shocks between financial

institutions is usually modelled by means of effective dynamics, which are only approximations of the true dynamics. In this thesis, I introduce a generalised stress testing model that captures a wide range of behavioural assumptions with regard to banks' liquidation dynamics under stress. The literature has proposed two alternative classes of liquidation dynamics in this regard, all of which are covered by my model: *threshold dynamics* (banks liquidate their investment portfolios only after they have defaulted), and *leverage targeting dynamics* (banks constantly rebalance their portfolios to maintain a target leverage ratio). I test the capability of the generalised model to predict actual bank defaults (and non-defaults) in the United States for the years 2008-10. I find that the model performs better than alternative benchmarks that do not account for the network of common asset holdings, irrespective of the assumed liquidation dynamics. I also show how the best performing liquidation dynamics depend on the combination of the initial shock level and the market impact parameter, on the cross-sectional variation in the market impact parameter, and on the number of asset liquidation rounds.

Finally, with respect to the third aspect, it is mandatory to define the types of financial institutions that the stress test covers. In this thesis, I study the impact of common asset holdings across different financial sectors on financial stability. In particular, I model systemic risk arising from indirect contagion between UK banks, UK open-ended investment funds and UK insurance companies. I show that performing a stress simulation that accounts for common asset holdings across multiple sectors results in systemic losses that are 47% larger than those obtained by summing the losses of sector-specific stress simulations. In other words, ignoring common asset holdings between different financial sectors may result in an underestimation of systemic risk.

Impact Statement

This thesis has a direct impact on the application of macroprudential stress testing. It provides central banks and regulators with methods to understand the consequences of taking a particular setup when conducting stress tests, and gives insights into how can stress tests be more accurately conducted. In particular, it suggests the best network reconstruction technique to use, when detailed interactions between financial institutions are not available. Moreover, it develops a structure approach to picking the desired contagion dynamics -based on the policymakers' specific criteria of interest- from a set of available options. Finally, it shows the importance of considering different financial sectors in stress tests, as ignoring common asset holdings between banks and non-banks can lead to a large underestimation of systemic risk.

Some works written in this thesis have been published in the Journal of Economic Dynamics and Control, some have been featured in the Deutsche Bundesbank Discussion Paper Series and in the Bank of England Staff Working Paper Series. They have been also presented in several regulatory meetings and conferences, such as in the Bank of England, the Deutsche Bundesbank, the Bank of Finland and the Bank of Indonesia Research Seminar.

Untuk Mama dan Ayah,
without whom I would be nothing.

Acknowledgements

It would be impossible for me to write this thesis without the generous help of many people. I would like to thank them below.

My first and deepest gratitude goes to my two supervisors, Dr. Fabio Caccioli and Dr. Daniel Fricke. Without their support, guidance and encouragement, this thesis would have never existed. At the beginning of the journey, many people said to me that I was lucky enough to be under Fabio's wing. Now I understand why. I also feel fortunate to have Daniel as my supervisor, as he not only inspires me to be a critical thinker but also encourages me to be a good writer. I will make good use of their teachings for the rest of my career, and I hope to continue this fruitful collaboration.

I gratefully acknowledge a PhD scholarship from the Indonesia Endowment Fund for Education (LPDP) that supported my studies.

I am very grateful to Prof. J. Doyne Farmer and Prof. Philip Treleaven, who kindly agreed to spend time reading and discussing this thesis.

Special thanks to Prof. Tomaso Aste and Dr. Giacomo Livan for their constructive comments during my first year and transfer viva. I am very proud to be part of the Financial Computing and Analytics Group for the last four years.

Some parts of this thesis benefited from the collaboration I had during my visit as a PhD intern at the Bank of England. A huge thanks to Dr. Gerardo Ferrara for being such a kind and considerate manager. Thank you for teaching me the beauty of *festina lente*. I also want to thank Nick Vause and Matt Roberts-Sklar for their valuable comments and generous guidance. Very warm thoughts to Simon Jurkatis, Nowrin Hossain, Miriam Kurtosiova and Arjun Mahalingam, who have been more

than colleagues but also friends to me.

I recall the time when I visited Lucca in May 2018. It was one of the best moments in my PhD life. I owe this feeling to the organisers and the participants of the Complexity72h workshop. Special thanks to my favourite people: Alberto, Domenico, Francesco, Mateusz and Valentina. I am also grateful to Dr. Giulio Cimini and Dr. Tiziano Squartini for their excellent comments on some works that I wrote in this thesis.

I would also like to send my warmest thoughts to those whom I am lucky enough to call my friends. First, I thank the friends of the WhatsApp group *Asian Women in Macro*: Vania, for her warm heart, Aruhan, for her jokes and Ivy, for her wisdom. I also thank Lili and Iis for their valuable *nasiha*. Big thanks to Indah and Elsa, who were there with me when I felt lost and alone, and to Abdi and Dadan, who always make me laugh.

Finally, my heartfelt gratitude goes to my family for their love and never-ending support: Dhila, my nonbiological sister, Ahmadi, my brother, and Ayah and Mama, to whom this thesis is dedicated.

Thank you,

Amy

London, May 2020

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"The duty of the man who investigates the writings of scientists, if learning the truth is his goal, is to make himself an enemy of all that he reads, and ... attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency."*
>
> – Ibn al-Haytham in *Doubts Concerning Ptolemy*

1.1 Motivation and Research Objective

The 2007-09 financial crisis has brought the interconnectedness of the financial system to light, and the interconnectedness of the financial system has been identified as an important source of systemic risk (May et al. (2008); Haldane and May (2011)). Accordingly, the regulatory framework has taken a more macroprudential perspective to maintain the stability of the system as a whole. For example, Basel III introduced capital surcharges for systemically important financial institutions. In this regard, the literature on network models of financial systemic risk has proliferated (see Hüser (2015); Glasserman and Young (2016); Caccioli et al. (2018) for the recent surveys), with a particular focus on the development of contagion network models (Neveu (2016); Aymanns et al. (2018)). More importantly, different systemic stress tests that account for interaction between financial institutions have been introduced in the literature (Furfine (2003); Eisenberg and Noe (2001); Gai and Kapadia (2010); Battiston et al. (2012); Huang et al. (2013); Caccioli et al.

(2014); Greenwood et al. (2015); Cont and Schaanning (2017)).

With the exception of Huang et al. (2013), relatively little work has been devoted to empirically testing the capability of any given stress testing model to predict financial systemic risk. Of course, generally speaking, stress tests are not a prediction tool. However, given that stress tests may be used as early-warning indicators, it is important to calibrate models in a way that they can capture the actual dynamics. This is especially important given that a multitude of models has been introduced in the literature. Moreover, considering the importance of stress testing in monitoring systemic risk, these questions are not of merely academic interest but of utmost practical importance. In fact, policymakers and regulators are well aware of the fact that the outcome of stress tests crucially hinge upon the underlying model, as pointed out for example by Niepmann and Stebunovs (2018) and Siemsen and Vilsmeier (2018).

This thesis is concerned with studying the predictive performance of systemic stress tests. The study in particular focuses on indirect contagion channel that is due to common assets holdings (overlapping portfolios) between financial institutions. This is different from direct contagion channel in interbank borrowing/lending networks that has received relatively more attention in the literature (Glasserman and Young (2015)). The indirect contagion dynamics is based on the idea of fire sales in asset markets: when a leveraged bank faces a loss, it may need to liquidate (part of) its assets. The corresponding market impact decreases the prices of the liquidated assets further, which creates a vicious circle where banks may need to sell even more assets in a falling market.

When a stress test methodology is developed, several modelling choices have to be made. For instance, what are the financial institutions that are going to be modelled? What is the network of interactions between these institutions? How do shocks propagate between interconnected institutions? These modelling choices affect the outcome of the stress test. Moreover, each assumption, in fact, represents an approximation to reality. For example, the network of financial linkages between financial institutions is often lacking, and one has to resort to network reconstruction

methods to infer the network from partial information. Similarly, the propagation of shocks between financial institutions is usually modelled by means of effective dynamics, which are only approximations of the true dynamics.

In this thesis, I will study the impact of modelling choices on the outcome of stress tests. I will focus on three particular aspects of stress tests related to indirect contagion: (i) the network of interaction, (ii) the dynamics of contagion, and (iii) the perimeter, or the types of institutions that are included in the stress test. In the different chapters, I will discuss how the choice that one makes, concerning each of these aspects, contributes to the accuracy.

With regard to the first aspect, stress tests require detailed data on interactions between individual financial institutions. However, it is difficult to collect such data in full and to make them readily available to researchers (e.g., due to data confidentiality), such that we generally do not have complete information about financial networks. For example, Haldane (2015) suggests that even among the world's largest banks, the collection of interbank exposure data is partial, and even regulators often do not have complete information (Glasserman and Young (2016)). Several data collection initiatives have been proposed, but granular interaction-specific data generally remain unavailable (Anand et al. (2018)). In this regard, different network reconstruction methods have been introduced in the literature (see, for example, Squartini et al. (2018)). It is therefore important to study the capability of these reconstruction methods to reproduce actual financial networks.

Concerning the second aspect, the literature has proposed different behavioural assumptions with regard to bank's liquidation dynamics under distress. For example, Huang et al. (2013) and Caccioli et al. (2014) assume that banks are passive during stress periods, and they will liquidate their assets only after they default. In contrast, Greenwood et al. (2015) assume that banks actively target their leverage ratio. It is not surprising that these dynamics may yield very different results. Therefore, properly defining and validating the dynamics of contagion is an important aspect of stress tests.

Finally, in terms of the third aspect, it is mandatory to define the perimeter,

in particular concerning the types of institutions that the stress test covers. Most of the literature focuses on quantifying systemic risk in a single financial sector, where most works have been devoted to the case of banks (see, for example, Caccioli et al. (2018) and Glasserman and Young (2016)). Recently, regulators have become concerned about the impact of non-banks (or more accurately non-bank financial intermediaries) on financial stability. As discussed in Fricke and Fricke (2020), this is due to the fact that the non-bank industry has grown enormously over the last decade, both in terms of size and importance. For example, the contribution of non-banks in the UK to the total assets of the UK financial system has increased by 13 percentage points since 2008, as it now accounts for almost 50% of the system (Baranova et al., 2019). Moreover, it has been empirically shown in Barucca et al. (2020) that banks and non-banks may have some portfolio overlap. In this respect, considering common asset holdings across financial sectors becomes a relevant aspect of future stress tests.

1.2 Thesis Structure

Figure 1.1 summarises the overall structure of this thesis. Firstly, I provide a litera-

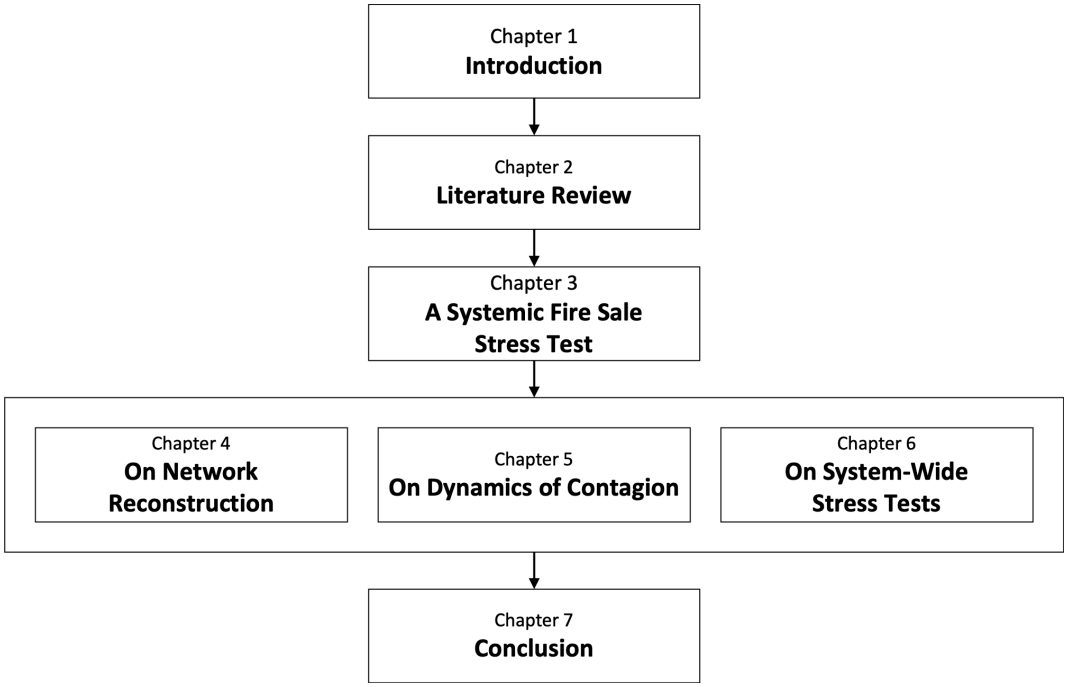


Figure 1.1: Thesis structure.

ture review in chapter 2. In chapter 3, entitled "A Systemic Fire Sale Stress Test", I present an overview of the thesis methodology. I note that the details of the methodology for each stress test aspect are provided in each subsequent chapter. In chapter 4, entitled "On Network Reconstruction", I conduct a horse race between different reconstruction methods, in terms of their abilities to reproduce actual bank-firm credit networks and observed systemic risk levels from partial information. Furthermore, in chapter 5, entitled "On Dynamics of Contagion", I propose a generalised contagion model that captures a wide range of behavioural assumptions with regard to bank's liquidation under distress. I then study the capability of the model to predict actual defaults (and non-defaults) of banks. In chapter 6, entitled "On System-Wide Stress Tests", I explore the importance of considering different financial sectors to stress tests. Finally, I conclude the results in chapter 7.

1.3 Thesis Contribution

Overall, I contribute to the literature on network models of systemic risk via the indirect contagion channel. As I discussed above, the focus of the literature has been on developing different stress testing models that may yield very different results. I take the literature forward by taking these models to the data and validating their output. In the following, I discuss the detailed contributions of the thesis on the three different aspects of stress tests.

1.3.1 On Network Reconstruction

Finding accurate reconstruction methods for financial networks from partial information is an important topic. However, most of the existing work of focuses on the case of interbank credit networks (Squartini et al. (2017); Gandy and Veraart (2017); Anand et al. (2018)). In this thesis, I use data on bank-firm credit relationships at different aggregation levels in Japan and conduct a horse race between different network reconstruction methods in terms of their ability to reproduce the actual credit networks. Interestingly, I find that there is no single "best" reconstruction method - it depends on the assumed criterion of interest. I then compare the different reconstruction methods in terms of their implied levels of systemic risk based on a

standard model of indirect contagion. To the best of my knowledge, this is the first work to conduct a horse race of bipartite network reconstruction methods in terms of their implied levels of systemic risk.[1] I find that the observed credit network displays relatively high levels of systemic risk compared with most reconstruction methods. In other words, many reconstruction methods tend to underestimate systemic risk. Lastly, I explore whether different policies can improve the robustness of the system.

Overall, this work contributes to different strands of literature: first, I add to the growing literature on reconstructing financial networks from partial information (Squartini et al. (2017); Gandy and Veraart (2017); Anand et al. (2018); Squartini et al. (2018)). For the case of bipartite networks, I am only aware of the works of Di Gangi et al. (2018) and Squartini et al. (2017). Given that most reconstruction methods have been designed for the case of unipartite credit networks, I adjust some of these methods to the case of bipartite networks. Second, I contribute to the literature on systemic risk assessment by performing stress tests both for the actual and reconstructed credit networks in Japan. Lastly, I contribute to the literature that explores the effects of aggregation on stress test results. For example, Hale et al. (2015) study the optimal aggregation level for stress testing models on macroeconomic variables, and they find that the aggregation level matters. I obtain a similar conclusion based on a completely different stress testing approach.

1.3.2 On Dynamics of Contagion

Two extreme types of banks' liquidation dynamics have been proposed in the literature: (i) *threshold dynamics* of Huang et al. (2013) and Caccioli et al. (2014), where banks liquidate their investment portfolios only after they have defaulted, and (ii) *leverage targeting dynamics* of Greenwood et al. (2015), where banks constantly rebalance their portfolios to maintain a target leverage ratio. I introduce a one-parameter family of non-linear liquidation functions that interpolates between these two extremes, and test the capability of these models to predict actual bank

[1] Some related papers for the case of unipartite interbank networks are Mistrulli (2011), Anand et al. (2015), and Gandy and Veraart (2017).

defaults (and survivals) in the United States for the years 2008–10.

My main findings are as follows: I show that the model performance depends on the type of shock being imposed (idiosyncratic versus systematic), and identify the two most relevant asset classes, for which the model has predictive power when these asset classes are exposed to an initial shock. In these cases, the model performs better than alternative benchmarks that do not account for the network of common asset holdings, irrespective of the assumed liquidation dynamics. I also discuss the fundamental differences between network models and statistical/econometric models in general, and argue that the former are more appealing to the application of systemic stress tests. Finally, I show how the best performing liquidation dynamics depend on the combination of the initial shock level and the market impact parameter, on the cross-sectional variation in the market impact parameter, and on the number of asset liquidation rounds.

In addition to the above-cited literature, my work mainly contributes to the following streams of the literature: first, my generalization of existing stress testing models captures a wide range of banks' asset liquidation behaviour in response to some initial shock. In this sense, my work is analogous to Bardoscia et al. (2016) who analyse counterparty risk within interbank networks. Second, my work adds to the literature on backtesting risk models. While backtesting microprudential risk models is now common practice among market practitioners (e.g., Philippon et al. (2017); Daníelsson (2011); Cavestany and Rodríguez (2015)), relatively little attention has been devoted to the case of macroprudential stress tests. My methodology allows us to compare the performance of different models and thus to identify the most accurate stress testing model, given exogenous parameters. This is similar to the approach of Huang et al. (2013), who test the predictive performance of the threshold model. In my work, I use the same methodology but allow for different liquidation dynamics and different combinations of the initial shock/market liquidity parameters. Lastly, my work is also related to the literature on reverse stress testing (e.g., Grigat and Caccioli (2017)). Fundamentally, reverse stress testing is concerned with identifying scenarios that will lead to a certain stress testing out-

come. In my analysis, the outcomes that I wish to match are individual defaults/non-defaults.

1.3.3 On System-Wide Stress Tests

In this thesis, I study the impact of common asset holdings across different financial sectors on financial stability. In particular, I model systemic risk arising from fire sale contagion between UK banks, UK open-ended investment funds and UK (both unit-linked and non unit-linked) insurance companies. I assume that fire sales are triggered by different responses to a financial shock: banks and non unit-linked insurers are subject to regulatory constraints, while funds and unit-linked insurers are obliged to meet investor redemptions. To this end, I build a network of common asset holdings from granular datasets of portfolio holdings across different sectors. I then conduct a systemic stress test under different initial shock scenarios and institutions' selling strategies. I find that performing a stress simulation that accounts for common asset holdings across multiple sectors results in fire sales losses that are 47% larger than those obtained by summing the losses of sector-specific stress simulations. This indicates that ignoring asset common asset holdings between different financial sectors may result in an underestimation of systemic risk.

My main contribution to the literature is the quantification of systemic risk on granular data of portfolio holdings of banks and non-banks. Most of the literature to date focuses on quantifying systemic risk in a single financial sector. For example, Duarte and Eisenbach (2015) and Cont and Schaanning (2017) look at systemic risk in the banking sector, while Baranova et al. (2017) and Fricke and Fricke (2020) model indirect contagion between investment funds. Moreover, this work is also related to the literature on portfolio similarity between different financial sectors. In this respect, my work is the closest to Barucca et al. (2020), who study the portfolio similarity between UK banks, UK insurance companies and European funds. However, while they consider holdings at the security issuer level (where each asset is identified by a LEI - Legal Entity Identifier), I instead study those at the ISIN level (where each asset is identified by using the International Securities Identification Number). Therefore, I consider more granular data of portfolio holdings.

1.4 Publications and Talks

The work I carried out in this thesis has also been described in several manuscripts and presented at several conferences:

1. Chapter 4 has been published as: ”Ramadiah, A., Caccioli, F., Fricke, D. (2020). Reconstructing and Stress Testing Credit Networks. Journal of Economic Dynamics and Control, 111. DOI 10.1016/j.jedc.2019.103817”.

    - It has been presented in the 2019 Royal Economic Society Symposium of Junior Researcher, 2018 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics and 3rd workshop on Statistical Physics for Financial & Economic Networks at NetSci 2018.

2. Chapter 5 has been featured in the Deutsche Bundesbank Discussion Paper Series as: ”Ramadiah, A., Fricke, D., Caccioli, F. (2020). Backtesting Macroprudential Stress Tests. Deutsche Bundesbank Discussion Paper No 45/2020”.

    - It has been presented in the Deutsche Bundesbank Research Seminar in August 2019, Workshop on Economic Science with Heterogeneous Interacting Agents 2019, Conference on Complex Systems 2018, Econophysics Colloquium 2018, and 2nd Financial Economics and Network Science Workshop.

3. Chapter 6 has been featured in the the Bank of England Staff Working Paper Series as: ”Caccioli, F., Ferrara, G., Ramadiah, A. (2020). Modelling Fire Sale Contagion across Banks and Non-Banks. Bank of England Staff Working Paper No. 878”.

    - It has been presented in the Bank of England Research Seminar in October 2019 and Complexity Meet Finance at NetSci 2020.

Chapter 2

Literature Review

2.1 Systemic Risk

It has been considered in the economic literature for a long time that financial systems are fragile, in a way that modest shocks may result in serious turmoil (Allen and Gale (1999)). In fact, many advanced models of a single bank run have been explored and developed in the banking literature (Bryant (1980); Diamond and Dybvig (1983); Chari (1988)). However, as pointed out in Battiston et al. (2016), it was not until the global financial crisis of 2007-09 that the phenomenon of financial contagion gained interest in the literature. This phenomenon is different from a single bank run, as the former generally involves multiple banks, and is stemming from the fact that distress in one institution may propagate to other institutions in the financial system. Moreover, it is considered to be one of the important components of systemic risk (Martínez-Jaramillo et al. (2010)).

Systemic risk has been formally defined in different ways in the literature. For example, Kaufman and Scott (2003) describe it as ”the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by comovements (correlation) among most or all the parts”. On the other hand, systemic risk is defined in the annual report of the European Central Bank (2004) as ”the risk that the inability of one institution to meet its obligations when due will cause other institutions to be unable to meet their obligations when due. Such a failure may cause significant liquidity or credit problems

and, as a result, could threaten the stability of or confidence in markets." Finally, Fouque and Langsam (2013) argue that "Systemic risk is the risk of a disruption of the markets ability to facilitate the flows of capital that results in the reduction in the growth of the global GDP".

Although the literature has not reached a consensus on the definition of systemic risk, it is illustrated above that a key element of systemic risk is the propagation of shocks in the financial system. This is based on the idea that shock is endogenously amplified through a complex network of exposures among banks, as pointed out in De Bandt et al. (2012). The existence of this complex interconnectedness has been empirically studied in the literature. For example, many have provided evidence that financial institutions are interconnected through credit relationships (Marotta et al. (2015); Fricke and Roukny (2020)), derivative contracts (Heise and Kühn (2012)), ownership relationships (Elliott et al. (2014); Vitali et al. (2011)) and common asset holdings (Duarte and Eisenbach (2015); Greenwood et al. (2015); Fricke (2019)).

As many contagion models have been relatively well developed outside economic literature (e.g., health and epidemic diseases literature), systemic risk has evolved to be an interdisciplinary study (De Bandt et al. (2012); Fouque and Langsam (2013)).[1] More importantly, since the interconnectedness can be conveniently modelled as a network (Hüser (2015)), the literature on network models of systemic risk has grown tremendously for the last decade (Aymanns et al. (2018); Caccioli et al. (2018);Glasserman and Young (2016)).

2.2 Systemic Stress Tests

According to International Monetary Fund (2012), stress testing is "a technique that measures the vulnerability of a portfolio, an institution, or an entire financial system under different hypothetical events or scenarios". Moreover, Baudino et al. (2018) define stress tests for banking sector as "simulation exercises conducted to assess

[1]There are also many studies of systemic risk in the economic literature whose approaches are typically based on macroeconomic modelling, game theory and finance. I refer the interested reader to Bisias et al. (2012) for a comprehensive survey on different measures of systemic risk.

the resilience to a hypothetical scenario of either a single bank or the system as a whole". Broadly speaking, stress tests can be grouped into two categories: microprudential and macroprudential. The former focus on the risk of an individual institution's failure, and can be completed by banks themselves or supervisors (Anderson et al. (2018)). An example of the microprudential stress tests is the exercise coordinated by the European Banking Authority.[2] Meanwhile, the latter are usually run by central banks and/or supervisory agencies to assess systemic risk (Anderson et al. (2018)), and are therefore also known as systemic stress tests. An example of these tests is the Euro area macroprudential top-down exercise conducted by the European Central Bank.[3] This thesis focuses on this second type of stress test.

With regard to the development of contagion network models (Aymanns et al. (2018); Neveu (2016)), various systemic stress tests based on different assumptions have been introduced in the literature (see Anderson (2016); Anderson et al. (2018); Dent et al. (2016) for a discussion on the development of stress tests). Once a systemic stress test has been developed, it is usually calibrated on empirical data sets to perform counterfactual simulations regarding the stability of the underlying financial networks. As pointed out in Upper (2011), stress tests have been used to assess the systemic risk level of various national interbank systems. For example, Puhr and Schmitz (2014) carry out a stress test exercise on the Austrian interbank exposure, while **?** apply the exercise on the Brazilian interbank network. Moreover, stress tests have been applied to measure fire sale spillovers due to portfolio overlap between financial institutions. For example, Duarte and Eisenbach (2015); Di Gangi et al. (2018); Huang et al. (2013) look at the contagion in the network of common asset holdings between U.S. banks, while Greenwood et al. (2015); Cappiello et al. (2015); Cont and Schaanning (2017) focus on the fire sale spillovers among European banks. The similar model has been also to the network of overlapping portfolios between U.S. funds (Fricke and Fricke (2020)) and European funds (Baranova et al. (2017)).

Alternatively, theoretical analyses are performed on stylized synthetic data sets

[2] https://eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing

[3] See Dees et al. (2017) for more details.

to study the effect of particular financial network topologies. For example, Gai and Kapadia (2010) study the contagion mechanism in interbank networks, and show that financial networks with arbitrary structure (Erdös-Rényi networks) are robust, but also fragile, at the same time. Moreover, Mastromatteo et al. (2012) observe that sparse Erdös-Rényi networks are more fragile compared to fully connected networks, while Roukny et al. (2013) argue that no single topology is always more robust than others. Furthermore, Caccioli et al. (2012) consider the model in Gai and Kapadia (2010), and study the effect of heterogeneous degree distributions, heterogeneous balance sheet size and degree correlations between banks. Banwo et al. (2016) then extend the work in Caccioli et al. (2012) to the case of common asset holdings networks.

However, with the exception of Huang et al. (2013), relatively little work has been devoted to empirically testing the capability of any given stress test model to predict actual bank defaults (and non-defaults). This is particularly important given that a multitude of models have been introduced in the literature, and many have been concerned with the fact that the outcomes of stress tests crucially hinge upon the underlying model. For example, Siemsen and Vilsmeier (2018) observe a huge dispersion in the outcomes of stress tests, after applying a number of plausible designs on the German banking sector dataset. Moreover, Niepmann and Stebunovs (2018) find that the flexibility observed in the 2014 and 2016 EBA stress tests was exploited by banks to manipulate the outcomes of stress tests. This thesis, therefore, focuses on studying the predictive performance of systemic stress tests, particularly on the indirect contagion channel that is due to common assets holdings (overlapping portfolios) between financial institutions.

2.3 Indirect Contagion in Financial Networks

The existing literature considers mainly two contagion channels of systemic risk in financial networks. The first one is the direct contagion channel in unipartite interbank borrowing/lending networks (Eisenberg and Noe (2001); Gai and Kapadia (2010); Furfine (2003); Battiston et al. (2012)), which has been the focus of the

literature over the last decade (see Hüser (2015), Glasserman and Young (2016) and Caccioli et al. (2018) for the recent surveys). The contagion in this channel occurs from banks' contractual obligations and is driven by a failure of a debtor to fully repay its creditors for the full amount of borrowed funds. In particular, the losses incurred in the creditors' balance sheet may, in turn, cause the creditors to default, which then makes the creditors unable to repay the creditors' creditors, and so on.

The second one is the indirect contagion channel due to common asset holdings (overlapping portfolios) between financial institutions. This contagion mechanism is based on the idea of fire sales in asset markets (see Shleifer and Vishny (2011) for a broad survey of the literature on fire sales), that is when leveraged investors suffer a decline in their investment portfolios, they often have to liquidate (parts of) their investments (Adrian and Shin (2010)). Such liquidations can have systemic effects, when asset sales are synchronized among many investors, potentially leading to fire sale contagion dynamics. Thus, investors who were unaffected by the original shock may have to sell assets due to the selling pressure of other investors. Empirical evidence suggests that fire sales occur in many different markets (see, e.g., Pulvino (1998) for real assets, Coval and Stafford (2007) for equities, and Ellul et al. (2011) for corporate bonds), which can result in contagious dynamics between asset classes (see, e.g., Manconi et al. (2012)).[4]

Although indirect contagion channel has received relatively less attention in the literature, it has been actually identified as an important source of systemic risk (Shleifer and Vishny (2011); Caccioli et al. (2014); Greenwood et al. (2015); Cont and Wagalath (2016); Gualdi et al. (2016); Lillo and Pirino (2015)).[5] Hence, understanding the predictive performance of systemic stress tests for this contagion channel is therefore important. As discussed above, this thesis focuses on three different aspects of stress tests related to indirect contagion. In the following, I

[4]Fire sales are also dangerous because they provide an incentive for banks to hoard liquidity, a behaviour that can potentially lead to a complete freeze of the financial system (Diamond and Rajan (2011); Gale and Yorulmazer (2013)).

[5]There are also several studies that combine both direct and indirect contagion channel. For example, Cifuentes et al. (2005) simulate a model that accounts for the interaction between the direct and indirect contagion channel, while Caccioli et al. (2015) and Poledna et al. (2018) empirically study the similar interaction for the case of Austrian and Mexican banks.

discuss the literature on each of these aspects.

2.4 Reconstructing Financial Networks

Stress tests require detailed data on interactions between individual financial institutions. However, we generally do not have such complete information about financial networks; thus a network reconstruction process is usually needed (see for example, Upper (2011)). The literature on network reconstruction is concerned with finding appropriate null models (i.e., network randomizations) that replicate certain features of the actual network, where most works have been devoted on the case of unipartite interbank networks (Anand et al. (2015); Squartini and Garlaschelli (2011); Squartini et al. (2017)). In this respect, several studies on the empirical test of unipartite network reconstruction methods have been conducted (Mistrulli (2011); BIS (2015); Anand et al. (2015, 2018); Mazzarisi and Lillo (2017); Gandy and Veraart (2017)).

In this thesis, I empirically test different network reconstruction methods for the case of the indirect contagion channel. As in Anand et al. (2018), I focus on the independent reconstruction of the static credit networks for each year. As such, I do not take into account the existence of previous bank-firm credit relationships and do not explicitly model time variation in the observed network topologies due to certain economic mechanisms. For example, additional models that explicitly take memory and/or preferential lending into account (De Masi and Gallegati (2012); Iori et al. (2015); Hatzopoulos et al. (2015)) could potentially improve network reconstruction. Here I focus exclusively on methods that preserve certain features of the observed network. This choice is justified by the fact that, among others, Hatzopoulos et al. (2015) find that preferential lending in interbank networks is largely driven by the degree distribution.

Existing reconstruction methods can be classified in terms of the inputs needed to reconstruct the network, the desired network features to preserve, and the outputs. To reconstruct a given interbank network, for example, all methods use the information of banks' aggregate borrowing and lending positions, respectively. In

this way, the total size of the system and the size of each individual market participant are expected to match the actual values. In addition, some methods also use the system's overall connectivity (Squartini et al. (2017)), while others use each bank's individual connectivity (Squartini and Garlaschelli (2011)). In terms of the desired network features, some methods focus on minimizing the total number of connections (Anand et al. (2015)), while others focus on minimizing the exposure with respect to each counterparty (Upper (2011)). Lastly, in terms of their outputs, some methods produce a single network for a given set of partial information, while others generate an ensemble of networks.

In chapter 4, I will look at four different network reconstruction methods that have been found to be of importance for unipartite financial networks (see, for example, Anand et al. (2015); Anand et al. (2018); BIS (2015); Gandy and Veraart (2017); Mazzarisi and Lillo (2017); Mistrulli (2011)). In the following, I discuss the development of the four reconstruction methods.

2.4.1 Details

Maximum Entropy (MaxEntropy). In the literature on financial networks, MaxEntropy is often considered as the standard approach to derive individual interbank liabilities in the absence of further information. It has been widely used to reconstruct interbank networks of different countries (see Upper (2011); Anand et al. (2015)). The main characteristic of MaxEntropy is that it generates fully connected networks, i.e., it assumes maximum diversification. Di Gangi et al. (2018) show that, in the case of bipartite networks, MaxEntropy implies that all market participants hold the exact same (market) portfolio.

Minimum Density (MinDensity). Second, I look at the Minimum Density approach (MinDensity) of Anand et al. (2015). This method was developed to acknowledge the fact that real financial networks tend to be sparse, in which case using MaxEntropy is rather problematic (Mistrulli (2011)). In a sense, MinDensity can be seen as the opposite extreme of MaxEntropy, given that it starts from the premise that establishing/maintaining links is costly, which is in line with the fact that most banking networks are sparse. As a result, banks do not spread their

borrowing and lending across the entire system, and MinDensity identifies the network that satisfies the total aggregate positions with the minimum number of links. This assumption is in line with the fact that relationship banking is of the utmost importance in most banking systems.

Configuration Models (CM). CMs are probably the most popular types of random graph models because they allow randomizing a given network while preserving its degree distribution. As such, CM can be quite restrictive. CMs have been previously explored in different fields, from sociology to biology (see Fosdick et al. (2018) for an overview), and several of them have been applied in financial network settings (Squartini and Garlaschelli (2011); Musmeci et al. (2013); Mastrandrea et al. (2014); Cimini et al. (2015b); Squartini et al. (2017)). I am aware only of one other application that applies the CM to bipartite financial networks (Squartini et al. (2017)). Two configuration models (CM1 and CM2) that I will consider in chapter 4 are based on Squartini and Garlaschelli (2011) and Squartini et al. (2017).

2.5 Modelling Dynamics of Contagion

Once the complete information of financial networks has been obtained, the next step is to determine the dynamics of contagion in the networks. With regard to banks' liquidation dynamics under distress, several network models of indirect contagion have been introduced in the literature. Here I provide a summary of the models' comparison based on the type of market impact function and whether it assumes some form of leverage targeting. As shown in Table 2.1, I first note that the model of Huang et al. (2013) uses a linear market impact and assumes that banks do not target their leverage. As in Huang et al. (2013), the model of Caccioli et al. (2014) also disregards leverage targeting, although the leverage targeting is incorporated in the extended version of the model. However, contrary to Huang et al. (2013), the model of Caccioli et al. (2014) uses a non-linear market impact. Similar to the extended version of Caccioli et al. (2014), the model of Greenwood et al. (2015) incorporates leverage targeting but assumes a linear market impact function. Another proposed model in the literature is the model of Cont and Schaanning

(2017), who do not include pure leverage targeting, but assume that banks have some regulatory constraint regarding their maximum leverage and banks will only liquidate when they exceed that maximum threshold. Another distinction between Cont and Schaanning (2017) and Greenwood et al. (2015) is that even though the model of Cont and Schaanning (2017) also assumes a linear market impact for small volumes, they use a non-linear impact function with heterogeneous price impacts for each asset class.

		Market impact	
		linear	*non-linear*
Leverage targeting	*not included*	Huang et al. (2013)	Caccioli et al. (2014)
	included with threshold		Cont and Schaanning (2017)
	included	Greenwood et al. (2015)	Caccioli et al. (2014) (extended)

Table 2.1: Comparison between different existing network models of contagion due to overlapping portfolio, based on the type of market impact function used and whether leverage targeting is included or not.

The above models have been considered in several studies of systemic risk. For example, Levy-Carciente et al. (2015) apply the model Huang et al. (2013) to the Venezuelan banking system, while Duarte and Eisenbach (2015) apply the model of Greenwood et al. (2015) to measure the systemic risk of U.S. banking system. Meanwhile, Coen et al. (2019) extend the banks' deleveraging model by taking additional regulatory constraints such as capital and liquidity coverage ratio constraint into account. In chapter 5, I will use and extend the model of Huang et al. (2013), Caccioli et al. (2014) and Greenwood et al. (2015).

2.6 Systemic Risk across Different Financial Sectors

In addition to network reconstruction and contagion dynamics, another important aspect of stress tests is to define the scope of the network. With regard to the indirect contagion channel, empirical studies suggest that fire sales can arise under different causes. For example, banks may be forced to deleverage in response to losses in their portfolios (Khandani and Lo (2011); Cont and Wagalath (2016)), while funds may be forced to sell their assets during stress periods to meet investor redemptions

(Coval and Stafford (2007)). Additionally, Ellul et al. (2011) suggest that companies may need to liquidate their assets to comply with regulatory constraints. Other empirical evidence also suggests that fire sale can occur in real assets (Pulvino (1998)). However, the literature on the indirect contagion channel in financial networks has only focused on quantifying systemic risk for one sector in isolation. In particular, as I discussed above, most works have been undertaken in the case of banks.

Lately, regulators have been concerned about the possibility of non-bank financial intermediaries to become another source of financial instability (European Central Bank (2014); International Monetary Fund (2015)). This is related to the fact that the asset management sector has drastically grown for the past decade, both in terms of its size and its importance. For example, the contribution of non-banks in the UK to the total assets of the UK financial system has increased by 13 percentage points since 2008, as it now accounts for almost 50% of the system (Baranova et al., 2019). In this respect, quantifying systemic risk in the non-banking sector has therefore recently gained interest in the literature. For example, Cetorelli et al. (2016), Fricke and Fricke (2020) and Baranova et al. (2017) model the indirect contagion in the network of common asset holdings between funds. Furthermore, Douglas et al. (2017) and Douglas and Roberts-Sklar (2018) study the impact of Solvency II regulation on the way that UK life insurers and pension funds adjust their portfolio in periods of distress.

While most of the above studies focused on a single financial sector in isolation, some works have looked at financial networks between different sectors. For example, Barucca et al. (2020) perform an empirical study on the financial network of common asset holdings between European investment funds, UK banks and UK insurance companies, and investigate the portfolio overlaps between those institutions. In chapter 6, I will study the systemic risk via indirect contagion in the financial network of common asset holdings between UK banks, UK investment funds and UK insurance companies.[6]

[6]There is also another stream of literature on modelling systemic risk with multiple interacting contagion, amplification channels and various financial sectors (see, for example, Aikman et al. (2019) and Farmer et al. (2020)). Most of the financial institutions in their models, however, are representative agents. In the following, I instead focus on empirical data on common asset holdings.

Chapter 3

A Systemic Fire Sale Stress Test

In the following, I describe the general methodology that I use to undertake the research. Note that the details of each aspect that I study in this thesis (network reconstruction, dynamics of contagion and system-wide stress tests) will be explained in the subsequent chapters.

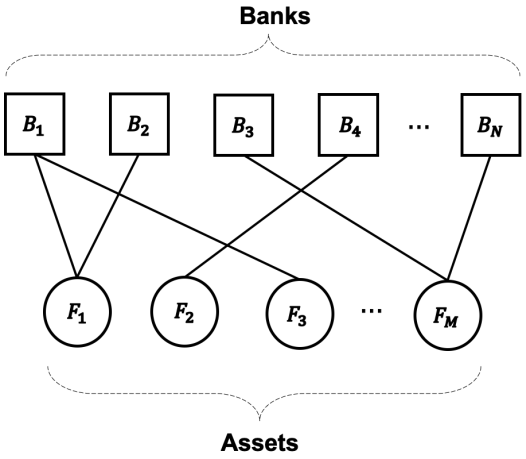
3.1 The Network of Common Asset Holdings

Let me start by defining the network of common asset holdings. The baseline network consists of two distinct sets of nodes, where the first set contains a total number of N nodes (financial institutions), and the second set contains a total of M nodes (financial assets). I illustrate the stylized network of common asset holdings in Figure 3.1. A link between an institution and an asset implies that the institution holds the asset in its portfolio. In general, the institution can be a bank, fund, or insurance company. In this chapter, I will mainly look at the case of banking networks.[1] Moreover, the asset can be a risky financial instrument (e.g., bond or equity), where a link will imply an investment relationship). Additionally, it can be a firm, where a link will correspond to a credit relationship. I should note that the network is bipartite, which means the absence of inter-institution and inter-asset links.

This financial network can be represented as a rectangular matrix of size $(N \times M)$, which I denote by $\mathbf{W}$. An element w_{ij} of this matrix represents the total value of asset j owned by bank i.[2] The value of w_{ij} can thus be seen as a measure of link

[1] I will discuss the case for non-bank financial intermediaries in chapter 6.

[2] I drop time subscripts in the following, but it should be clear that matrix $\mathbf{W}$ changes over time.



	F_1	F_2	F_3	...	F_M
B_1	1	0	1	...	0
B_2	1	0	0	...	0
B_3	0	0	0	...	1
B_4	0	1	0	...	0
...	...	...	...	...	...
B_N	0	0	0	...	1

Figure 3.1: The left panel shows the stylised network of common asset holdings. A link between a bank and an asset will imply that the bank holds the asset in its portfolio. The right panel illustrates the corresponding rectangular matrix that represents this network.

intensity. The total investment in the network can be calculated as

$$v = \sum_i \sum_j w_{ij}.$$

For what follows, it is also useful to define the *strength* of banks as their total assets:

$$s_i^B = \sum_j w_{ij},$$

and the *strength* of assets as their total value of shares:

$$s_j^F = \sum_i w_{ij}.$$

I also define the binary adjacency matrix, $\bar{W}$, where each element $\bar{w}_{ij} = 1$ if $w_{ij} > 0$ and zero otherwise. From the binary network matrix, I can calculate the total number of links existing in the network:

$$m = \sum_i \sum_j \bar{w}_{ij}.$$

This value can be also represented as a proportion of the number of possible connections that could exist in the network:

$$\text{Density(W)} = \frac{m}{N \times M}.$$

In addition, I define the *degrees* of banks and assets as their corresponding number of connections:

$$k_i^B = \sum_j \bar{w}_{ij}$$

and

$$k_j^F = \sum_i \bar{w}_{ij}$$

for bank i and asset j, respectively.

3.1.1 Network Measures

In the following, I describe several quantities to characterize the topology of the network (e.g., assortativity, clustering coefficient and nestedness). Furthermore, I discuss several measures of portfolio similarity.

3.1.1.1 Network Characteristics

Assortativity is the tendency of banks to connect to assets with similar degree, and vice versa. I define assortativity, r, as the Pearson correlation coefficient of the degrees of connected banks and assets. Note that r lies in the range $[-1,1]$ in which positive value indicates an assortative network, while negative value denotes a disassortative network. A network is said to be assortative when high degree banks (low degree banks) are connected to other high degree assets (low degree assets) on average. Meanwhile, a network is said to be disassortative when high degree banks (low degree banks) are connected to other low(er) degree assets (high(er) degree assets) on average.

The *clustering coefficient* measures the degree to which nodes in a network tend to form clusters. In a unipartite network, it is usually defined as the number of observed triangles (three closed connected nodes) relative to the maximum possible number of triangles. Since my network is bipartite, links can only exist between

different sets of nodes (banks and assets), thus triangles can not be formed. Therefore, following Zhang et al. (2008), I consider squares instead of triangles as the basic cycle here, such that the local clustering coefficient is defined as the ratio between the number of observed squares relative to the maximum possible number of squares,

$$C_{mn}(i) = \frac{q_{imn}}{(k_m - 1 - q_{imn}) + (k_n - 1 - q_{imn}) + q_{imn}}$$

where m and n are a pair of neighbours of node i (see Figure 3.2 for an illustration), and q_{imn} is the number of squares which include these three nodes.

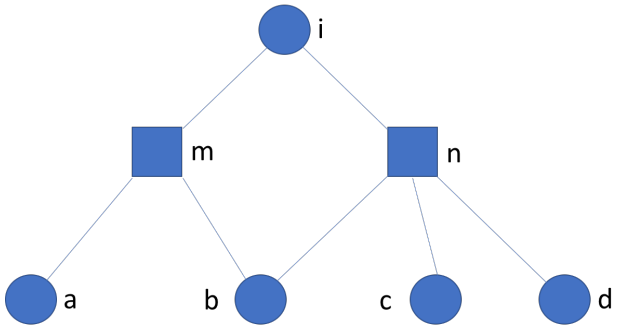


Figure 3.2: Illustration of calculating the observed and the possible squares in a bipartite network (Zhang et al. (2008)). In this figure, m and n are a pair of neighbors of node i. Here I observe 1 square cycle (q_{imn} = 1) that consists of node *imbn*, and 4 possible squares (*iman*, *imbn*, *incm*, *indm*).

Let $C^{row}(i)$ and $C^{col}(i)$ be the average $C_{mn}(i)$ of node i across all possible combination of its pairs of neighbors m and n, I then calculate the global clustering coefficient as

$$C = \frac{1}{N+M}\left(\sum_{i=1}^{N} C^{row}(i) + \sum_{i=1}^{M} C^{col}(i)\right),$$

which ranges between $[0,1]$; higher values indicate a more clustered network, and a value of 1 corresponds to a perfectly clustered network. Put simply, in my case, higher clustering will indicate that banks tend to cluster their investments on the same assets, or equivalently, assets tend to be held by the same banks.

Lastly, *nestedness* quantifies the degree to which low-degree banks (assets) tend to be connected with a subset of assets (banks) that are connected to high-degree nodes. I follow Almeida-Neto et al. (2008) and use NODF (*Nestedness*

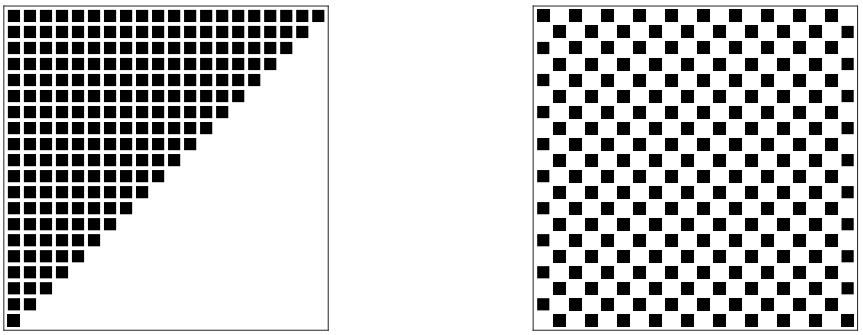
*metric based on **O**verlap and **D**ecreasing **F**ill*) as my measure of nestedness

$$\text{NODF} = \frac{\sum_{ij} G_{ij}^{row} + \sum_{ij} G_{ij}^{col}}{N(N-1)/2 + M(M-1)/2},$$

where

$$G_{ij}^{row} = \begin{cases} 0 & \text{if } k_i \leq k_j \\ \sum_{d=1}^{M} \mathbb{I}\{\bar{w}_{id} = 1 \text{ AND } \bar{w}_{jd} = 1\}/\min(k_i, k_j) & \text{otherwise.} \end{cases}$$

is the paired overlap of rows i and j, which is simply the fraction of 1's (which denotes to the existence of a link) present in both rows i and j. A similar term G_{ij}^{col} is used to compute the percentage of paired overlap of columns i and j. NODF lies in the range $[0,1]$; higher values correspond to higher nestedness, and a value of 1 indicates a perfectly nested network. I provide the illustration of networks with high and low nestedness in Figure 3.3.



(a) High nestedness. **(b)** Low nestedness.

Figure 3.3: Illustration of networks with high and low nestedness degree.

3.1.2 Portfolio similarity

To quantify the *portfolio similarity* between banks i and k in the network, I use two different types of measure. The first one is based on the binary matrix $\bar{W}$:

$$\text{BinSimilarity}_{i,k} = \sum_{j} \bar{w}_{ij} \bar{w}_{kj}, \tag{3.1}$$

which shows the number of commonly held assets between banks i and k. The other measure is the cosine similarity between the portfolio weights:

$$\text{WeightedSimilarity}_{ik} = \frac{\sum_j w_{ij} w_{kj}}{\sqrt{\sum_j w_{ij}^2} \times \sqrt{\sum_j w_{kj}^2}}. \tag{3.2}$$

Note that BinSimilarity ranges between $[0, \infty]$, while WeightedSimilarity ranges between $[0, 1]$. For both measures, higher values correspond to more similar portfolios.

In addition to these measures, I am also interested to look at the portfolio similarity between a bank and all other banks in the network. Following Fricke (2019), I define the average portfolio overlap as:

$$\text{MeanBinSimilarity}_i = \frac{1}{N-1} \sum_k \text{BinSimilarity}_{ik} \tag{3.3}$$

for the binary similarity case, and

$$\text{MeanWeightedSimilarity}_i = \frac{1}{N-1} \sum_k \text{WeightedSimilarity}_{ik}. \tag{3.4}$$

for the weighted one.

3.2 The Balance Sheet of a Bank

I previously discussed in section 3.1 the fact that each bank i in the network holds a portfolio of assets $\{w_{i1}, ..., w_{iM}\} \geq 0$. In the following, I will consider these quantities to depend on time t, so that the total assets of bank i at time t is:

$$A_i^t = \sum_{j=1}^{M} w_{ij}^t.$$

Each bank i is financed with a mix of equity E_i^t and liabilities L_i^t (see Figure 3.4) with a balance-sheet identity:

$$A_i^t \equiv E_i^t + L_i^t,$$

and bank i does not default at time t if:

$$A_i^t \geq L_i^t.$$

The *leverage* of bank i is defined as the ratio of its total assets to its equity,

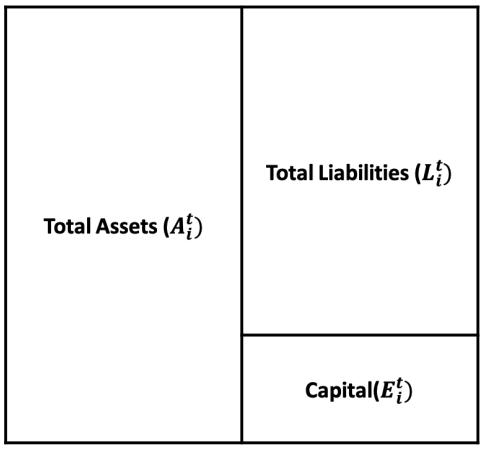


Figure 3.4: The stylized balance sheet of bank i. The left panel is the asset side, while the right is the liability side of the balance sheet.

$$\lambda_i^t = \frac{A_i^t}{E_i^t}. \tag{3.5}$$

Bank's i initial leverage, λ_i^0, will be of particular interest for the *leverage targeting model*.

3.3 The Dynamics of Contagion

The steps of the contagion algorithm can be summarized as follows:

1. An initial shock is imposed on the value of some asset(s).
2. Banks update the value of their portfolios, and compute the total assets that they need to sell.
3. Banks liquidate their assets by maintaining their portfolio weight (pro-rata liquidation), or by selling their most liquid assets first (waterfall liquidation).
4. Asset liquidations generate price impact, so the value of an asset is recomputed depending on the volume of the asset that has been liquidated.
5. Back to step 2 (optional).

Let us discuss the above steps in detail.

3.3.1 Step 1: Initial Shock

Suppose I initially impose a shock on asset j by reducing its value to a fraction $p \in [0,1]$ of its original value, where larger (smaller) values of p correspond to smaller (larger) shocks. Assuming that a given bank i holds asset j in its portfolio, the initial shock leads to a reduction of its total assets:

$$A_i^1 = A_i^0 - w_{ij}^0(1-p_j^0).$$

The absolute *return* on assets of bank i at time 1 is therefore:

$$\phi_i^1 = \frac{|A_i^1 - A_i^0|}{A_i^0}, \tag{3.6}$$

and bank i's updated equity becomes:

$$E_i^1 = E_i^0 - \phi_i^1 A_i^0,$$

while its liabilities remain unchanged, $L_i^1 = L_i^0$. Note that bank i's updated leverage ratio reads as:

$$\lambda_i^1 = \frac{(1-\phi_i^1)A_i^0}{(1-\phi_i^1)E_i^0 - \phi_i^1 L_i^0} \geq \lambda_i^0. \tag{3.7}$$

Hence, in response to a drop in asset values, leverage will mechanically increase when liabilities remain fixed (Adrian and Shin (2010)). Finally, bank i has not defaulted if

$$\phi_i^1 \leq \frac{1}{\lambda_i^0}. \tag{3.8}$$

3.3.2 Step 2: Response to the Shock (Asset Liquidation)

The change in bank i's total assets may trigger it to liquidate some of its assets: if the devaluation is sufficiently large to make bank i default, that is when Equation 3.8 is no longer satisfied, bank i liquidates all of its remaining assets. On the other hand, if it has not yet defaulted, bank i may sell part of its assets.

Let us start by looking at the *leverage targeting model* of Greenwood et al. (2015). In their model, bank i has a fixed leverage target: λ_i^0. As shown in Eq. 3.7, the initial shock will increase the bank's leverage, and the bank will have to liquidate a fraction of its assets to maintain its original leverage target λ_i^0. In this case, the total volume of the liquidated asset (in monetary units) is therefore specified as:

$$\Pi_i^1 = A_i^1 \left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1}\right). \tag{3.9}$$

To accommodate for the case where bank i defaults and needs to sell all of its remaining assets, I modify the function Π_i^1 in Equation 3.9 as follows:

$$\Pi_i^1 = A_i^1 \min\left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1}, 1\right).$$

In contrast to the leverage targeting model, the *threshold model* (Huang et al. (2013); Caccioli et al. (2014)) assumes that bank i will only liquidate assets when it

defaults.[3] Let us introduce G_i^1, such that

$$\Pi_i^1 = G_i^1(\phi) A_i^1 \min\left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1}, 1\right), \tag{3.10}$$

where in the leverage targeting model I have:

$$G_i^1(\phi) = 1, \tag{3.11}$$

while in the threshold model:

$$G_i^1(\phi) = \begin{cases} 0, & \text{if } \phi_i^1 \leq \frac{1}{\lambda_i^0}, \\ 1, & \text{otherwise.} \end{cases} \tag{3.12}$$

I consider the threshold model in chapter 4, while I use the leverage targeting model in chapter 6. Moreover, I propose $G_i^1(\phi)$ in a way that one can interpolate between the leverage targeting and threshold dynamics inchapter 5.

3.3.3 Step 3: Liquidation Strategy

Once the total amount to be liquidated has been computed, different liquidation strategies could be used. In what follows, there are two scenarios that I consider. The first one is the pro-rata liquidation, where banks maintain their portfolio weights constant over time. The second scenario is the waterfall liquidation, where banks liquidate assets in order of their liquidity starting from the most liquid ones. I illustrate the comparison between these two approaches in Figure 3.5.[4]

3.3.3.1 Vertical assets liquidation

Some studies have suggested that the pro-rata liquidation is more favourable for financial institutions during distress (Jiang et al. (2017); Schaanning (2016)). This is based on the idea that institutions want to preserve the liquidity of their portfolios. Suppose that π_{ij} is the amount of asset j that bank i choose to liquidate. In the case

[3] Note that the extended version of the model in Caccioli et al. (2014) incorporates the leverage targeting dynamics.

[4] In a setting with homogeneous price impacts, however, the two approaches are equivalent.

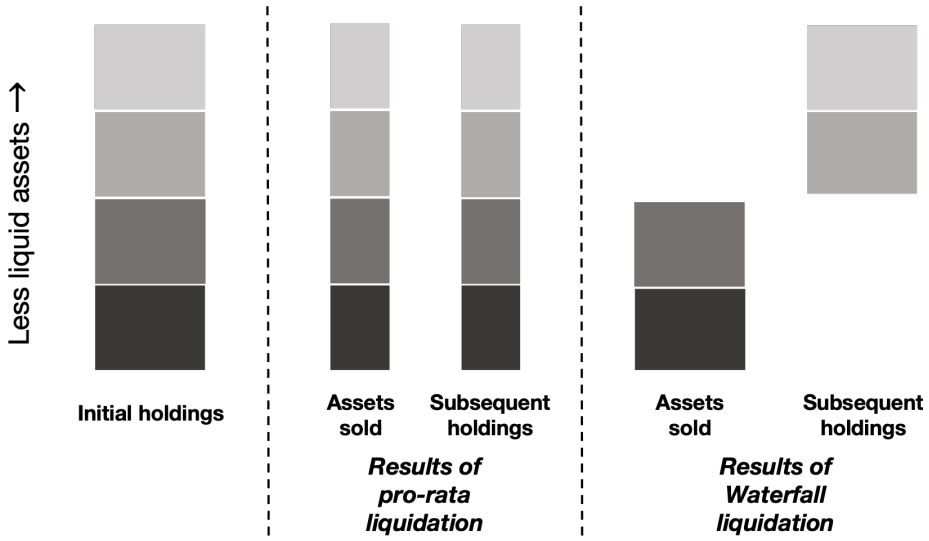


Figure 3.5: Illustration of pro-rata and waterfall liquidations.

of the pro-rata liquidation, I will have:

$$\pi_{ij}^{1} = \frac{w_{ij}^{1}}{A_i^{1}} \Pi_i^{1}.$$

3.3.3.2 Horizontal assets liquidation

In the case of the waterfall liquidation, I assume that bank i liquidates its assets sequentially according to the following order:

$$\text{sort}\left\{\delta_1 \geq \cdots \geq \delta_M\right\},$$

where δ_j is the market depth of asset j, that is the measure of j's liquidity to sustain relatively large transactions without impacting its price. This approach is supported by Chernenko and Sunderam (2016), who provide empirical evidence that funds use cash holdings, rather than transacting in equities and bonds, to meet investor redemptions.

3.3.4 Step 4: Fire Sales Generate Price Impact

In line with empirical evidence, this additional selling pressure generates market impact, which leads to a further devaluation of the assets.[5] Let β_j^1 be the amount of asset j that has been liquidated,

$$\beta_j^1 = \sum_{i=1}^{N} \pi_{ij}^1.$$

The relative price change of asset j, $\frac{\Delta S_j^1}{S_j^1}$, is equal to:

$$\frac{\Delta S_j^1}{S_j^1} = -\Psi_j(\beta_j^1),$$

where Ψ_j is the market impact function of asset j. The functional form of Ψ_j varies across different stress testing models (see Table 2.1 in Section 2 for the classification of existing models), and I refer the interested reader to Cont and Schaanning (2017) for a comprehensive discussion of this topic. In this thesis, I consider two different types of market impact function: the linear and non-linear one.

3.3.4.1 Linear market impact

I consider the market impact function that has been previously used in Huang et al. (2013). In particular, an asset's price depends linearly on the fraction of shares that have been liquidated up to that time relative to the total volume held in the system:

$$\Psi_j(\beta_j^1) = \alpha \frac{\beta_j^1}{\sum_i w_{ij}^0}, \tag{3.13}$$

where α is the parameter that reflects the market reaction to asset liquidations. More illiquid assets should have higher values of α: a value of $\alpha = 0$ corresponds to an infinitely liquid asset whose price does not change in response to asset liquidations. This will be the case for cash holdings, for example. On the other hand, a value of $\alpha > 0$ corresponds to a less liquid asset whose price reacts to assets liquidations.

[5] I should note that banks can also "liquidate" cash (banks use cash to repay existing debt), but without any market impact.

I note that choosing a linear impact function can be seen as more conservative, in the sense that I tend to overestimate the resulting price impact of a given asset liquidation.

3.3.4.2 Non-linear market impact

As in Cont and Schaanning (2017), I use the following market impact function:

$$\Psi_j(\beta_j^1) = 0.5 \times \left(1 - \exp(-\frac{\beta_j^1}{0.5 \times \delta_j})\right),$$

where δ_j is the market depth of asset j. In Figure 3.6, I illustrate the comparison between the linear and non-linear market impact function. I see from the figure that the non-linear function is concave and it is similar to the linear specification for small volumes of liquidation. Additionally, unlike in the linear case, the price impact in the non-linear function is capped at 50% of the initial asset price.

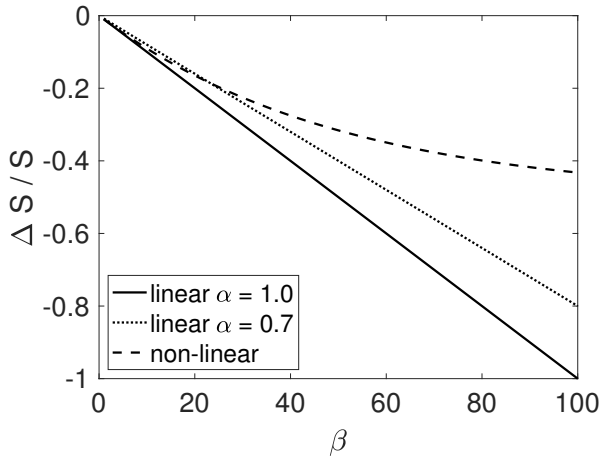


Figure 3.6: Illustration of the comparison between the linear (for $\alpha = 1$ and $\alpha = 0.7$) and non-linear market impact function, with $\sum_i w_{ij}^0 = 100$ and $\delta_j = 100$.

3.3.5 Step 4: Back to Step 2 (Optional)

I have described the first complete iteration round of the contagion dynamics, which results in asset liquidations generating price impact. Following the update in asset prices, banks will experience another decline in their total assets, which will again trigger them to liquidate (part of) their assets. Therefore, one can consider further rounds of liquidation in the model by going back to step 2.

3.4 Overview of the Specific Methodologies

I previously discussed the thesis methodology in general. In the following, I briefly describe the specific methodologies that I use in subsequent chapters to study the different aspects of stress testing. Table 3.1 summarises these methodologies. As shown in the table, I consider different types of financial networks, and assume different bank's response, liquidation strategy and market impact function throughout the thesis. For example, I consider bank-firm credit relationships in Chapter 4, and banks' portfolio holdings in Chapter 5. Moreover, I assume that banks would only liquidate their assets after default in Chapter 4, and banks may choose whether to target their leverage ratios in Chapter 5. The details of these methodologies will be discussed in each corresponding chapter.

	Financial networks	**Response to the shock**	**Liquidation strategy**	**Market impact function**
Chapter 4: On Network Reconstruction	Bank-firm credit relationships in Japan	Banks do not target their leverage ratio	Pro-rata liquidation	Linear and homogeneous
Chapter 5: On Dynamics of Contagion	Portfolio holdings of US commercial banks	Banks may target their leverage ratio	Pro-rata liquidation	Linear and homogeneous
Chapter 6: On System-Wide Stress Tests	Portfolio holdings of UK banks and non-banks	Banks target their leverage ratio	Pro-rata and waterfall liquidation	Non-linear and heterogeneous

Table 3.1: Overview of the specific methodologies that I consider in subsequent chapters to study the different aspect of stress tests.

Chapter 4

On Network Reconstruction

4.1 Introduction

In this chapter, I focus on reconstructing and stress testing bipartite credit networks using detailed micro-data on bank-firm credit interactions in Japan for the period 1980 - 2010. I explore the performance of several network reconstruction methods at different aggregation levels along two different dimensions. First, I look at their capability to reproduce the topological features of the observed credit networks. In particular, different reconstruction methods require different amounts of information as inputs, and I aim to understand how adding such information affects a method's performance, since one would expect that methods that take more information into account should be able to reproduce the network more accurately. Interestingly, I find that this is not always the case. Overall, there is no single "best" reconstruction method - it depends on the assumed criterion of interest.

I then test each method's ability to reproduce observed levels of systemic risk. For this purpose, I use the stress test model of Huang et al. (2013) and apply it to the actual and the reconstructed credit networks. My main findings are as follows: first, I identify a significantly negative time trend for the observed systemic risk levels of the Japanese banking system, suggesting that the system has become less vulnerable to systemic asset liquidations over time. Second, in many instances, the actual credit networks display the highest levels of systemic risk compared to the reconstructed networks, at least for the most disaggregated bank-firm interac-

tions. In other words, many reconstruction methods tend to underestimate systemic risk. This is remarkable given that the reconstruction methods under study here can generate completely different network architectures; for example, the MaxEntropy (MinDensity) approach yields a maximally (minimally) connected credit network. Moreover, I find that the network aggregation level can affect the performance of different reconstruction methods.

Lastly, given that the observed credit networks tend to display relatively high levels of systemic risk compared to most reconstruction methods, I explore different policies (such as merging or breaking-up banks, or leverage caps) in order to improve the robustness of the system. My main finding is that no single policy can reduce the systemic risk level of the actual network to that of the most stable reconstruction method. Nevertheless, I find that leverage caps and bank mergers could improve the robustness of the network. This finding is driven by the fact that the largest banks in my sample tend to be less leveraged. Therefore, merging those banks results in a very large but moderately leveraged bank, which is less likely to spread shocks through the system.

The remainder of this chapter is structured as follows: Section 4.2 defines the methodology that I use, including the credit network at different aggregation levels and the dataset. In section 4.3, I explore the performance of network reconstruction methods in terms of their ability to match the observed credit network topology. Additionally, I look at the capability of each method to reproduce the observed levels of systemic risk. In section 4.4, I analyse different policy measures in order to improve the robustness of the system. Section 4.5 summarises the main findings and concludes.

4.2 Methodology

In the following, I discuss the specific methodology and dataset that I use in this chapter.

4.2.1 Network at Different Aggregation Levels

The most granular data (disaggregated level) that I consider in this chapter is the credit interaction network between banks and firms. This network corresponds to $\mathbf{W}$ of dimension $(N \times M)$ as I previously defined in chapter 3. Following Fricke and Roukny (2020), I also look at an aggregated version of the credit (bank-industry) network, which I denote by $\mathbf{W^I}$. In this case, the second set of nodes is defined based on firms' industry affiliations, with a total number of M^I industries. I can represent firms' industry affiliations using a new matrix $\mathbf{A}$ of dimension $(M \times M^I)$, where $A_{jk} = 1$ if firm j is affiliated with industry k.[1] Given this, $\mathbf{W^I}$ can be obtained by multiplying $\mathbf{W}$ with $\mathbf{A}$. In line with the definitions for the original bank-firm credit network, I can define the same network indicators (strength and degree sequences, respectively) for the aggregated network.

Note that an important reason for also exploring the aggregated networks is that (at least some rough) information on banks' investments in different industries/asset classes should be more easily available than detailed microdata on asset-specific investments. From this perspective, the analyses based on the aggregated networks are likely to be most relevant for researchers that have only relatively coarse information on banks' asset portfolios.

Finally, I consider an intermediate level in which I apply the network reconstruction methods at the disaggregated level (bank-firm) and then aggregate the network according to firms' observed industry affiliations (thus giving us a different bank-industry credit network). This particular aggregation level is of interest in the case when there is sufficient data to perform network reconstruction at a more granular level (e.g., firm level), but the network needs to be analyzed at a more aggregated level (e.g., sector level), for instance because of confidentiality issues that prevent reporting results associated with individual institutions. I denote the intermediate aggregation level as $\mathbf{W} \rightarrow \mathbf{W^I}$ and calculate the same network indicators also as for the other levels. I summarize the three different aggregation levels in

[1] In my dataset, each firm is only affiliated with its major industry. In principle, one could allow for multiple industry affiliations, in which case a_{jk} would represent the fraction of firm j's sales in industry j.

Table 4.1.

Aggregation level	**Network reconstruction**	**Systemic risk analysis**
Disaggregated	disaggregated	disaggregated
Aggregated	aggregated	aggregated
Intermediate	disaggregated	aggregated

Table 4.1: Summary of the three different aggregation levels. At the intermediate level, I perform the network reconstruction at the disaggregated data, and conduct the systemic risk analysis at the aggregated version of that reconstructed network.

4.2.2 Data

I use historical data on bank-firm credit interactions in Japan from the Nikkei NEEDS database for the period 1980 - 2013.[2] The database provides extensive accounting and loan information for all listed companies in Japan, and since 1996 it also covers firms traded in the JASDAQ (OTC) market. The dataset contains information on firms outstanding loan volumes from each lender at the end of the firms fiscal year, based on survey data (compiled by Nikkei Media Marketing, Inc.). I use the sum of short- and long-term borrowing in everything that follows. Table 4.2 shows some summary statistics in terms of the size and connectivity of the credit network at different aggregation levels over time.[3] Given that my analyses are computationally intensive, I restrict myself to the years of data as shown in the first column of Table 4.2.[4]

In Table 4.2, I present several basic network characteristics of my dataset. Specifically, I show the assortativity, the clustering coefficient, and the nestedness. I find that the networks are generally disassortative, both at the disaggregated level and the aggregated level. This means that low-degree banks and low-degree firms rarely interact with each other. Table 4.2 shows that the networks are clustered at the aggregated level but not at the disaggregated one. Table 4.2 shows that all networks are nested at both aggregation levels, suggesting a strong overlap of Japanese

[2] See `https://www.nikkeieu.com/needs/needs_data.html` for details.

[3] A detailed explanation of the dataset, summary statistics, and a brief history of the Japanese financial system can be found in Fricke and Roukny (2020).

[4] Given that bank-firm interactions are highly persistent, the structure of the credit network is quite stable over time.

Panel A - Disaggregated

Year	Size	v ($\times 10^{13}$)	Density	$\bar{k}^B$	$\bar{k}^F$	r	C ($\times 10^{-2}$)	NODF
1980	151 × 1386	3.395	0.093	128.377	13.986	-0.299	0.272	0.441
1985	148 × 1443	4.350	0.088	127.770	13.105	-0.290	0.251	0.437
1990	148 × 1443	6.249	0.081	125.762	12.236	-0.306	0.218	0.427
1995	145 × 1734	7.031	0.081	140.938	11.785	-0.302	0.212	0.444
1996	147 × 2523	7.525	0.070	175.782	10.242	-0.292	0.141	0.406
2000	135 × 2607	5.987	0.061	160.304	8.301	-0.273	0.091	0.387
2005	123 × 2569	2.469	0.042	109.423	5.184	-0.272	0.029	0.322
2010	116 × 2296	2.814	0.042	96.474	4.874	-0.215	0.028	0.359

Panel B - Aggregated

Year	Size	v ($\times 10^{13}$)	Density	$\bar{k}^B$	$\bar{k}^I$	r	C	NODF
1980	151 × 33	3.395	0.516	17.033	77.939	-0.336	0.192	0.824
1985	148 × 33	4.350	0.500	16.507	74.030	-0.344	0.181	0.823
1990	151 × 33	6.250	0.498	16.424	75.152	-0.351	0.181	0.810
1995	145 × 33	7.031	0.518	17.090	75.091	-0.341	0.195	0.834
1996	147 × 34	7.526	0.536	18.238	78.853	-0.344	0.206	0.852
2000	135 × 34	5.987	0.508	17.260	68.529	-0.349	0.177	0.839
2005	123 × 34	2.470	0.488	16.585	60.000	-0.340	0.151	0.822
2010	116 × 34	2.814	0.461	15.664	53.441	-0.330	0.134	0.819

Table 4.2: Properties of the credit networks at different aggregation levels over time. Panel A shows the properties of $\mathbf{W}$. Panel B shows the properties of $\mathbf{W^I}$. $\bar{k}^B$ and $\bar{k}^{F(I)}$ correspond to the average degree of banks and firms (industries) respectively. As defined in the main text, r denotes the assortativity, C denotes the clustering coefficient, and NODF denotes the nestedness.

banks' loan portfolios (see Fricke and Roukny (2020)).[5]

4.2.3 Network Reconstruction Methods

I consider four network reconstruction methods that have been found to be of importance for unipartite financial networks, as I previously discussed in section 2.4. In particular, firstly, I look at the well-known method of Maximum Entropy (MaxEntropy). Second, I look at the Minimum Density approach (MinDensity). Lastly, I use two different versions of the popular configuration model (CM). While the description of each method has been presented in section 2.4, Table 4.3 provides more technical details of my implementation of the four methods. Moreover, Table 4.4 summarizes the differences in terms of the required inputs and the outputs.

[5]Note that these values cannot be used to assess the significance of nestedness. For this, one would have to compare them with what would be expected at random, i.e., using different null models. This is not the aim of this chapter, but the results in Table 4.5 suggest that the actual credit networks indeed tend to show higher NODF values than their random counterparts.

Null model	Required information	Definition and remarks
Configuration Model 1 (CM1)	k^B, $k^F(k^I)$, s^B, and $s^F(s^I)$ sequences	Generates ensemble of networks. Link allocation: based on the approach of Squartini and Garlaschelli (2011), but adjusted for bipartite network. The probability of link existence between every two nodes in the network, $p_{ij} = \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j}$, is calculated by solving: $\sum_j \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j} = k_i \quad \forall i, \quad \sum_i \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j} = k_j \quad \forall j.$ for θ and γ. Weight is allocated using RAS.
Configuration Model 2 (CM2)	s^B and $s^F(s^I)$ sequences and m	Generates ensemble of networks. Using fitness model. Link allocation: based on the approach of Squartini et al. (2017). The probability of link existence between every two nodes in the network, $p_{ij} = \frac{z s_i s_j}{1+z s_i s_j}$, is calculated by solving $\sum_i \sum_j \frac{z s_i s_j}{1+z s_i s_j} = m$ for z. Weight is allocated using RAS.
Maximum Entropy (MaxEntropy)	s^B and $s^F(s^I)$ sequences	Generates one single network. Produces completely connected network. Economic interpretation: each node is as diversified as possible. The approach is formulated as: $\min_x \sum_{i=1}^N \sum_{j=1}^M X_{ij} \log\left(\frac{X_{ij}}{\bar{X}_{ij}}\right)$, s.t. $\sum_{j=1}^M X_{ij} = s_i \quad \forall i = 1,2,...,N$ $\sum_{i=1}^N X_{ij} = s_j \quad \forall j = 1,2,...,M$
Minimum Density (MinDensity)	s^B and $s^F(s^I)$ sequences	Generates ensemble of networks due to multiple possible solutions. Economic interpretation: each node is as specialized as possible. Based on the approach of Anand et al. (2015), but adjusted for the case of bipartite networks. The approach is formulated as: $\min_x c \sum_{i=1}^N \sum_{j=1}^M 1_{\{X_{ij}>0\}}$, s.t. $\sum_{j=1}^M X_{ij} = s_i \quad \forall i = 1,2,...,N$ $\sum_{i=1}^N X_{ij} = s_j \quad \forall j = 1,2,...,M$ where the integer function 1 equals one if bank i lends to firm j, and zero otherwise.

Table 4.3: Summary of different network reconstruction methods used in this chapter.

Method	Input			Output	
	Aggregate positions	Total links	Degree sequence	Single	Ensemble
CM1	v	v	v		v
CM2	v	v			v
MaxEntropy	v			v	
MinDensity	v				v

Table 4.4: Summary classification of the network reconstruction methods based on the input and the output.

I should stress that, in contrast to MaxEntropy and MinDensity, both CMs produce binary instead of weighted networks.[6] After obtaining a randomized adjacency matrix, I need to distribute the observed credit volumes across links. There are different approaches for this (see Table A.1 in the Appendices for an overview), but in the following I use the standard RAS algorithm of Blien and Graef (1998).[7] It should be also clear that CM1 requires the most detailed information as inputs (followed by CM2), while MaxEntropy and MinDensity require only the strength sequences. In particular, CM1 requires the degree sequences of all nodes as additional inputs, thus preserving the exact degree distributions. CM2 preserves the degree distribution as well, but it only requires the total number of links as an additional input. Hence, CM2 needs less detailed information compared to CM1. Furthermore, CM1, CM2 and MinDensity can produce an ensemble of networks,[8] while MaxEntropy generates one single output for any particular input.

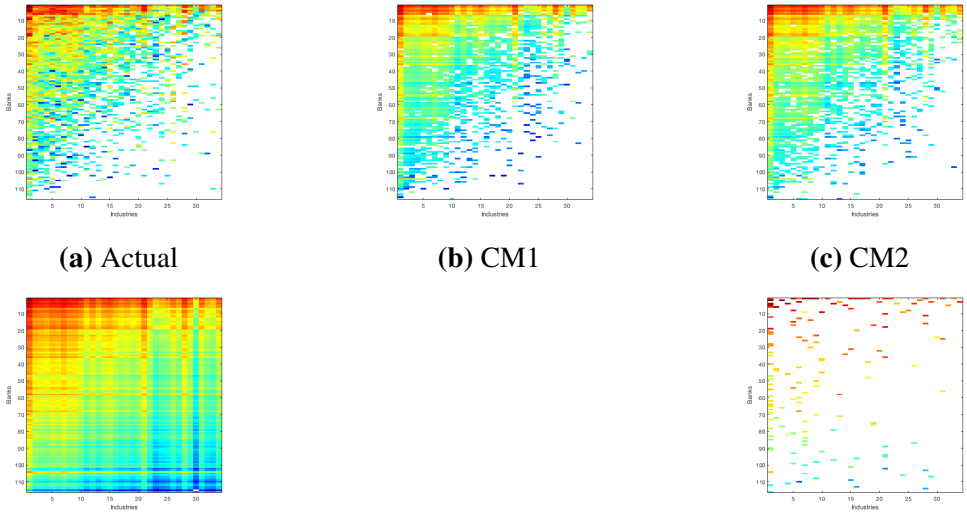
4.2.3.1 Illustration

In order to provide some intuition for the typical outputs of each method, Figure 4.1 shows the weighted version of the actual aggregated credit network (log-transformed) for the year 2010 and one realization of each corresponding null

[6]The original model of Squartini et al. (2017), where CM2 is based on, generates weighted networks. However, here I only consider part of their method to produce binary networks. This part of their method is based on the work of Saracco et al. (2015) where the formalism for the fitness bipartite is first introduced for the world trade web.

[7]The RAS algorithm generally performed best in my analysis (in terms of the corresponding L_1-error), but I also experimented with the other weight allocation methods mentioned in Table A.1 in the Appendices. The results are qualitatively similar to what is shown here. Details available upon request from the authors.

[8]Since the MinDensity-algorithm may yield multiple solutions, I treat this algorithm as an ensemble method.



(a) Actual **(b)** CM1 **(c)** CM2

(d) MaxEntropy **(e)** MinDensity

Figure 4.1: Weighted credit network bank-industry in 2010 and one realization for each of the four reconstruction methods. Data are log transformed. Warmer colours indicate stronger links, and white dots correspond to the absence of a link.

model. Warmer colours denote stronger relationships, and white dots correspond to the absence of a link. It becomes clear that different reconstruction methods can generate very different network architectures - for example, MaxEntropy produces a fully connected credit network while MinDensity yields a highly compartmentalized and sparse network. In this specific case, MinDensity needs less than 5% of the total links in the actual network to distribute the weight (the actual density is around 46%).[9] The two CMs, on the other hand, tend to produce networks that are visually much closer to the actual one. As such, it is natural to expect that these will perform well.

4.2.4 Defining Relevant Dimension of Comparison

In this section, I define the different dimensions along which I will compare the actual credit networks and the reconstruction methods.

[9] In my specific case, the bank-firm networks are sparse as well (see Table 4.5). On the other hand, the aggregated bank-industry networks are dense, such that MinDensity is likely to have difficulties in replicating the aggregated networks.

4.2.4.1 Network Characteristics

To understand how similar the statistics of the reconstructed networks are to the actual networks, I compare their density, average degree, assortativity, clustering and nestedness (as defined in chapter 3) at the different aggregation levels.

4.2.4.2 Allocation of Links and Weights

In addition to comparing specific network properties, I also look at the performance of each method in terms both of placing links and distributing weights correctly, respectively. In the following, I formally define the network similarity measures for the bank-firm credit network. In line with these definitions, I can define similar measures for the bank-industry credit network.

Link Allocation. In order to understand the ability of a method to reproduce correct links in the network, I calculate the values of Accuracy, Sensitivity, and Specificity. I define the *Accuracy* of a given reconstructed network as

$$\text{Accuracy} = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} (\mathbb{I}\{\bar{w}_{ij} = 0 \textit{ and } \bar{w}'_{ij} = 0\} + \mathbb{I}\{\bar{w} = 1 \textit{ and } \bar{w}'_{ij} = 1\}), \quad (4.1)$$

where $\bar{w}'_{ij}$ equals 1 if there is a link between nodes i and j in the reconstructed network of a given method. The indicator function, $\mathbb{I}(\omega)$, takes the value 1 if the event ω happens, and 0 otherwise. Put simply, Accuracy tells us the total number of links and non-links that are allocated correctly, relative to the size of the network.

Sensitivity

$$\text{Sensitivity} = \frac{1}{m} \sum_{i=1}^{N} \sum_{j=1}^{M} (\mathbb{I}\{\bar{w}_{ij} = 1 \textit{ and } \bar{w}'_{ij} = 1\}), \quad (4.2)$$

measures the number of actual links correctly allocated.

Lastly, *Specificity*

$$\text{Specificity} = \frac{1}{M \times N - m} \sum_{i=1}^{M} \sum_{j=1}^{N} (\mathbb{I}\{\bar{w}_{ij} = 0 \textit{ and } \bar{w}'_{ij} = 0\}, \quad (4.3)$$

measures the number of non-existing links correctly allocated. These three mea-

sures take values in the range $[0,1]$, with higher values corresponding to greater similarity.

Weight Allocation. I am also interested in quantifying the ability of each null model to reproduce the observed link weights in the credit network. For this purpose, I use three different measures: L_1-error, root-mean-square deviation (RMSE) and cosine similarity (Cos-Sim). L_1-*error* is defined as

$$L_1 = \sum_{i=1}^{N} |s_i^{B\prime} - s_i^B| + \sum_{j=1}^{M} |s_j^{F\prime} - s_j^F| \tag{4.4}$$

which allows us to understand how well the reconstructed network is able to satisfy the aggregate positions, which is the total borrowing (lending) of banks (firms/industries), in the actual network. As mentioned previously in Table 4.3 and Table 4.4, all null models are expected to reproduce the actual aggregate positions. Therefore, L_1-error measures the degree to which those constraints have been satisfied by a given null model. In everything that follows, I scale the L_1-error by the average lending volume of banks in the actual network.

Additionally, I calculate *RMSE* which is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(w'_{i,j} - w_{i,j})^2}{N \times M}}, \tag{4.5}$$

where $w'_{i,j}$ is the allocated credit volume of bank i to firm j in a given reconstructed network. In everything that follows, I scale RMSE by the average exposure of a link in the actual network which makes values comparable over time. Finally, I define *Cosine Similarity* as

$$Cos - Sim = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} w'_{i,j} w_{i,j}}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} w'^2_{i,j}}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} w^2_{i,j}}}. \tag{4.6}$$

to quantify deviations in the weight allocation across all links in the network.

I note that L_1-error and RMSE have values in the range [0,∞] with lower values corresponding to greater similarity. Meanwhile, Cos-Sim has values in the range

[0,1] and higher values correspond to greater similarity.

4.2.5 Measuring Systemic Risk

In order to measure the systemic risk, I consider the *threshold model* of Huang et al. (2013). As described in section 3.3, the model assumes that banks does not target their leverage ratios and would only liquidate their assets after default.[10]

Moreover, I consider a linear market impact function as in Equation 3.13. Note that α is a homogeneous (identical across assets) market impact parameter: a value $\alpha = 0$ corresponds to an extremely liquid asset, that is when any sales would not alter the market value of the asset, while $\alpha = 1$ corresponds to an extremely illiquid asset, where sales could potentially push the market price down to 0.[11] I will show results for different values of p and α. In the following, I mainly focus on a specific range of parameters. In particular, I consider loans to be relatively illiquid and therefore focus on the upper range of the market impact parameter ($\alpha \in [0.6, 1]$).[12] Moreover, in line with previous studies on price-mediated contagion, I consider relatively small values of the initial shock ($p \in [0.6, 1]$).[13]

I perform the above exercise for each industry j. At the aggregated level, for each iteration I shock one node (industry), while for the disaggregated level I shock all the nodes (firms) that belong to the same industry. To quantify the impact of a shock on industry j, I first define the *default rate*

$$\text{default}_j = \frac{N_j^{default}}{N} \tag{4.7}$$

as the ratio between the number of failed banks to the total number of active banks

[10] For the purpose of finding out how the systemic risk analysis might vary if leverage targeting model (as in Greenwood et al. (2015)) and leverage targeting with threshold model (as in Cont and Schaanning (2017)) are used, I also performed the same exercise with these other models. I find that the rank ordering of the different methods are generally consistent with those presented in the main text. See Appendices for more details.

[11] Among other things, the liquidity of a loan might be dependent on its remaining maturity. I leave a detailed calibration of the market impact parameter for future work.

[12] For $\alpha = 0.7$, the asset price drops by 7% when 10% of the asset is liquidated; for $\alpha = 1$, the price drops by 10% when 10% of the asset is liquidated.

[13] Greenwood et al. (2015) consider a 50% write-off on GIIPS debts, while Cont and Schaanning (2017) gradually increase the shock from 0% to 20%.

in the network. I then define the *probability of default*

$$P_d = \frac{\sum_{j=1}^{M^I} \text{default}_j}{M^I} \tag{4.8}$$

as the average of default_j across all industries. This is my systemic risk measure, and in the following I use the terms systemic risk and P_d interchangeably. Finally, for a given reconstructed network W', I also define *relative difference* between the actual P_d^W and the null model $P_d^{W'}$ as

$$D_r = \frac{P_d^W - P_d^{W'}}{P_d^W}. \tag{4.9}$$

A positive (negative) value of D_r indicates that a given null model underestimates (overestimates) the actual P_d.

4.3 Horse Racing Results

4.3.1 First Dimension: Topological Features

In this section, I show results of a horse race between the different reconstruction methods. For each year under study and each null model, I generate 100 network realizations for each aggregation level. I then calculate the average of each of the characteristics mentioned previously in 3.1.1.1. For the sake of brevity and also illustrative purposes, in the following I only show results for the year 2010, but the results are qualitatively similar for other years and do not affect my main conclusions.

The main results for the three aggregation levels can be found in Tables 4.5 (network statistics) and 4.6 (link/weight similarity). In all cases, the *best* method for each statistic is highlighted using the $\star$ symbol. Let us briefly describe the results for the different aggregation levels.

4.3.1.1 Disaggregated Level

At the disaggregated level (bank-firm), the top panel of Table 4.5 shows that the two CMs tend to reproduce the features of the actual network reasonably well: the

Network characteristics						
Disaggregated	Density	$\bar{k}^B$	$\bar{k}^F$	r	C ($\times 10^{-2}$)	NODF
$\mathbf{W}(116 \times 2296)$	0.042	96.474	4.874	-0.215	0.028	0.359
CM1	0.042	96.601	4.881	⋆-0.205	⋆0.028	⋆0.366
CM2	⋆0.042	⋆96.510	⋆4.876	-0.321	0.062	0.254
MaxEntropy	1.000	2296	116	NaN	1.000	0.000
MinDensity	0.009	20.789	1.050	-0.125	0.000	0.009

Network characteristics						
Aggregated	Density	$\bar{k}^B$	$\bar{k}^I$	r	C	NODF
$\mathbf{W^I}(116 \times 34)$	0.461	15.664	53.441	-0.330	0.134	0.819
CM1	⋆0.460	⋆15.649	⋆53.392	⋆-0.370	⋆0.136	⋆0.821
CM2	0.461	15.683	53.507	-0.248	0.131	0.704
MaxEntropy	1.000	34.000	116.000	NaN	1.000	0.000
MinDensity	0.038	1.285	4.385	-0.224	0.000	0.044

Network characteristics						
Intermediate	Density	$\bar{k}^B$	$\bar{k}^{F\to I}$	r	C	NODF
$\mathbf{W} \to \mathbf{W^I}$	0.461	15.664	53.441	-0.330	0.134	0.819
CM1	⋆0.482	⋆16.395	⋆55.936	⋆-0.308	⋆0.152	⋆0.798
CM2	0.493	16.771	57.218	-0.289	0.175	0.769
MaxEntropy	1.000	34.000	116	NaN	1.000	0.000
MinDensity	0.178	6.055	20.658	-0.329	0.019	0.442

Table 4.5: Comparison of the statistics between the actual credit network for year 2010 and the reconstructed networks for different aggregation levels. $\bar{k}^B$ and $\bar{k}^F$ ($\bar{k}^I$) correspond to the average degree, r denotes the assortativity, C indicates the clustering coefficient and NODF denotes the nestedness of the network. I highlight the best reconstruction method for a given statistic (the value closest to the actual network) using the ⋆ symbol.

density, average degree, assortativity, clustering, and nestedness are always quite similar to the actual values. On the other hand, MaxEntropy and MinDensity perform rather poorly: for example, in terms of density MaxEntropy (MinDensity) produce much higher (lower) values.

The results for link allocation and weight distribution (top panel of Table 4.6), are broadly in line with those for the network characteristics: again the two CMs perform relatively well across the different measures. In this case, however, the results are not always consistent. For example, MinDensity achieves the highest Accuracy and the lowest L_1-error, but shows the worst Sensitivity, RMSE, and Cos-Sim. On the other hand, MaxEntropy yields the worst Accuracy but the best RMSE and Cos-Sim. Not surprisingly, MaxEntropy achieves the maximum Sensitivity

	Link similarity			Weight similarity		
Disaggregated	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	0.941	0.304	0.969	4.511	18.674	0.442
CM2	0.936	0.241	0.967	2.706	13.850	0.633
MaxEntropy	0.042	⋆1.000	0.000	0.000	⋆13.038	⋆0.681
MinDensity	⋆0.955	0.071	⋆0.994	⋆0.000	27.896	0.278

	Link similarity			Weight similarity		
Aggregated	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	⋆0.781	0.762	0.798	0.015	⋆2.527	⋆0.915
CM2	0.711	0.687	0.732	0.018	2.555	0.914
MaxEntropy	0.461	⋆1.000	0.000	⋆0.000	2.572	0.914
MinDensity	0.558	0.061	⋆0.982	0.000	8.607	0.532

	Link similarity			Weight similarity		
Intermediate	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	⋆0.767	0.771	0.764	4.511	2.675	0.905
CM2	0.738	0.750	0.726	2.706	⋆2.530	⋆0.915
MaxEntropy	0.461	⋆1.000	0.000	0.000	2.572	0.914
MinDensity	0.668	0.333	⋆0.954	⋆0.000	3.676	0.836

Table 4.6: Link and weight similarity of the reconstruction methods with the actual credit network in 2010 for different aggregation levels. Accuracy, sensitivity, specificity and cosine similarity lie in the range [0,1] and higher values correspond to higher similarity. L_1-error and RMSE lie in the range [0,∞] with smaller values corresponding to greater similarity. I highlight the best reconstruction method for a given statistic (the value closest to the actual network) using the ⋆ symbol.

simply because it predicts a fully connected network.

I should also mention that both CM1 and CM2 generate relatively large L_1-errors, indicating that they do not manage to perfectly allocate the aggregate positions. This is because CM1 and CM2 preserve the degree sequences only in expectation, such that specific realizations can lead to some low-degree nodes being inactive (or unconnected).[14]

4.3.1.2 Aggregated Level

Similar to the previous results, the center panel of Table 4.5 shows that the two CMs tend to reproduce the observed network characteristics reasonably well at the aggregated (bank-industry) level. In this particular case, CM1 consistently performs

[14] I also experimented with a minimum threshold in terms of active nodes' degrees, but observe a similar issue.

best. For link/weight similarity, the results are also comparable (center panel of Table 4.6), except for Sensitivity and Specificity which are again dominated by MaxEntropy and MinDensity, respectively.

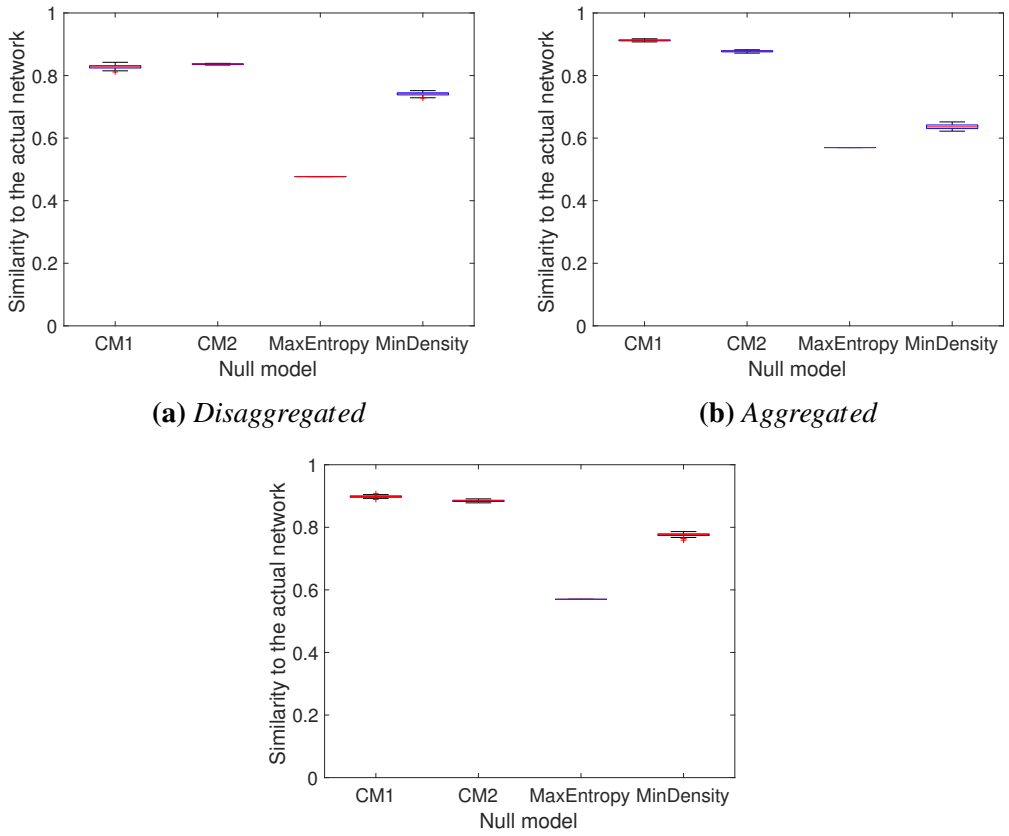
4.3.1.3 Intermediate Level

Lastly, the two bottom panels of Tables 4.5 and 4.6 show the results for the intermediate aggregation level, where I construct synthetic networks for the disaggregated (bank-firm) level and then aggregate these to the industry level using firms' observed industry affiliations. Overall, the statistics shown here are very similar to those at the aggregated level (with the exception of the L_1-error, which is close to the value at the disaggregated level), with CM1 performing best for the network statistics and the Accuracy.

4.3.1.4 Summary and discussion: topological features

Previous studies on the reconstruction of interbank networks (e.g., Anand et al. (2018)) suggest that the best reconstruction method depends on the type of network characteristics of interest. My findings support this conclusion and extend it to the case of bipartite networks of banks and firms. I see, for example, that if I focus the horse race on the number of non-existent links in the adjacency matrix that are correctly estimated (Specificity), MinDensity which produces sparse networks is clearly the winner. However, when I look at the number of links correctly estimated (Sensitivity), MaxEntropy, which generates a fully connected network, outperforms all other methods.

Given that my comparison is based on multiple network statistics, Figure 4.2 summarizes the results by combining statistics for the individual features. For this purpose, I normalize each measure in Table 4.5 (network characteristics) and Table 4.6 (link similarity and weight distribution) such that each of them ranges between 0 and 1, where the value of 1 (0) indicates that the reconstructed network and the actual network are identical (completely different) in terms of these features. I compute the metrics for each synthetic network, and take the average of these quantities. Figure 3 shows a standard box-plot of the metrics of each reconstruction



(a) *Disaggregated* **(b)** *Aggregated*

(c) *Intermediate*

Figure 4.2: Standard box-plots of the (normalized) similarity measures between the actual network and different network reconstruction methods (averaged over many realizations). For each reconstruction method, I consider each of the measures in the three categories of similarity under study: network characteristics, link similarity, and weight distribution. I normalize each of these measures such that a value of 1 (0) indicates that the reconstructed network and the actual network are identical (completely different) in terms of these features. I compute the metrics for each synthetic network, and take the average of these quantities.

method that are averaged over the realizations of the synthetic networks.[15] Overall, I find that the two CMs consistently perform the best, followed by MinDensity and MaxEntropy. I also note that, in general, CM1 and CM2 succeed in reproducing the topological structure related to the heterogeneity of links in the actual network (e.g., assortativity). This is important since heterogeneity plays an important role for systemic risk in financial networks (Iori et al. (2006), Banwo et al. (2016)).

Since CM1 and CM2 require more information relative to the other methods

[15]Note that I ignore the average degree (since it is redundant with density) and assortativity (since it is not defined for MaxEntropy) from the calculation.

(degree sequence and total degree, respectively), it seems clear that adding such information improves the performance of the reconstruction methods (see also Mastrandrea et al. (2014) and Cimini et al. (2015a)). This finding is in line with Gandy and Veraart (2016), who suggest that using the information on aggregate positions only is not sufficient to reconstruct certain topological properties of the network. Overall, it seems reassuring that, despite the fact that CM1 requires more information than CM2, both methods generate very similar networks (in some cases CM2 even outperforms CM1). This indicates that the degree distribution of the network might indeed, to a certain extent, be inferred without the full knowledge of the degree sequence. An obvious follow-up question is to what extent CM2 would still perform well if I treated the overall density as a free parameter. I leave this question for future research.

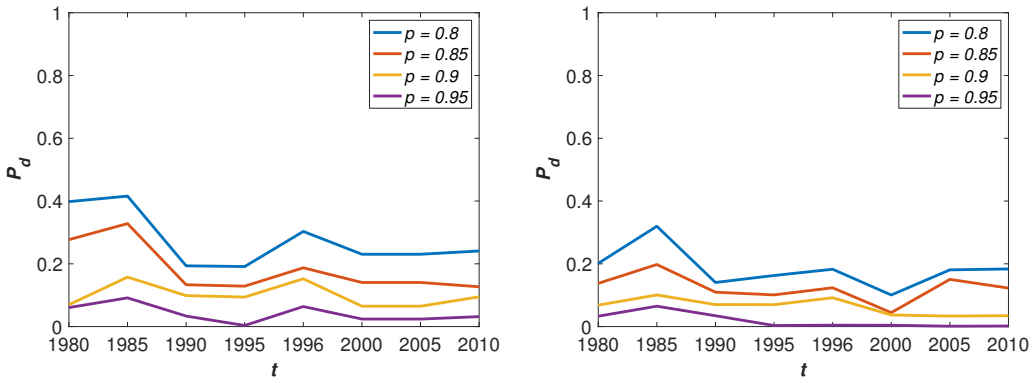
4.3.2 Second Dimension: Systemic Risk Analysis

One of the main reasons why regulators and policymakers are interested in reconstructing financial networks from partial information is because of their potential contribution to financial instability. Therefore, exploring how well different methods are able to reconstruct the observed networks is only the first step. The next step is to compare how well the different network reconstruction methods are able to replicate the levels of systemic risk of the actual credit networks. Clearly, this analysis is not independent from the results of the previous section, in the sense that one would expect a method that closely reproduces the actual networks to also yield similar systemic risk levels. To the best of my knowledge, however, such an exercise has not been performed for the case of bipartite financial networks.

4.3.2.1 Time dynamics of systemic risk

Before going into the details regarding the different reconstruction methods, I first quantify the level of systemic risk, P_d, over time. Figure 4.3 plots the P_d over time, both for the disaggregated (left panel) and the aggregated level (right panel), respectively. As a benchmark, I use a market impact parameter $\alpha = 0.7$ and different values of the initial shock p. The plots in Figure 4.3 suggest that P_d is smaller in

2010 compared with the values earlier in the sample. In other words, in many instances the level of systemic risk appears to have been reduced over time.



(a) W: Disaggregated level **(b) $\mathbf{W^I}$** Aggregated level

Figure 4.3: P_d over time for the disaggregated (left panel) and the aggregated level (right panel). I use $\alpha = 0.7$ and various values of p.

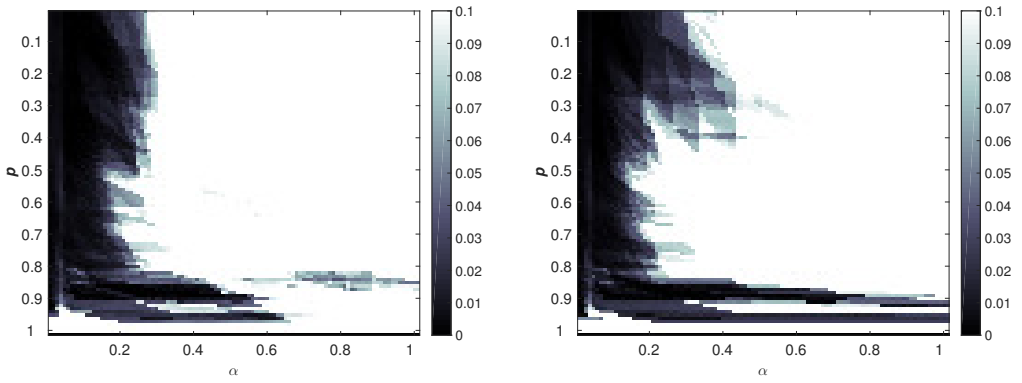
I also test for a significant trend in P_d for different values of α and p.[16] I then plot the corresponding p-value of the estimated trend as a heatmap in Figure 4.4, where darker colours correspond to smaller p-values (i.e., significance) of the estimated trends. The Figure shows that I obtain a significant trend for most values of p (except for very large values) whenever α is relatively small.[17]

4.3.2.2 Results on horse racing different methods

I now turn to a detailed analysis of the different null models and their implied levels of systemic risk. As before, I focus my presentation on the results for one particular year of data, namely 2010, but the results shown here are again broadly consistent over time. I will show three sets of results: first, Figures 4.5-4.9 show heatmaps of D_r for all possible combinations of p and α. Recall that D_r is the relative difference between the systemic risk levels of the actual and the reconstructed network, where a positive (negative) value of D_r indicates that the latter underestimates (overestimates) the former (see Equation 4.9). Second, Figure 4.6 allows us to take a closer

[16] Technically, for a given combination of α and p, I regress the resulting P_d on a constant and a time variable (year).

[17] For relatively large values of α the absence of a time trend in P_d is easily explained by the fact that in these cases all banks will tend to default in every single year. Hence, P_d will be roughly constant over time.



(a) W **(b) $\mathbf{W^I}$**

Figure 4.4: Trend analysis. p-value of regression analysis of P_d against a constant and a time variable (year), for different combinations of p and α. Darker colour denotes a smaller p-value.

look at the systemic risk levels, P_d, for a specific choices of α as a function of p in the range $p \in [0.6, 1]$. Third, to illustrate that my findings are robust over time, Figure 4.7 shows the P_d's over time for specific choices of α and p.

As for the network reconstruction part in section 4.2.3, I briefly discuss the results separately for the three different aggregation levels. Table 4.7 then summarizes these results.

4.3.2.3 Disaggregated level

Figure 4.5 shows that the actual network tends to be the riskiest, because all null models underestimate the actual P_d for most values of p and α. I observe that this underestimation is consistent for the range of parameters that I consider here (small initial shock and high market impact), which is shown by an area inside the black dashed line in Figure 4.5. Moreover, all null models overestimate the actual P_d only in a small region of the parameter space, for example when $p = 0.9$ (small initial shock) and $\alpha = 0.1$ (small market impact). Figure 4.5 also shows that the magnitude of the underestimation gets larger as α increases.

Panel (a) in Figure 4.6 shows the performance of the null models for a specific value of $\alpha = 0.7$ as a function of specific values of $p \in [0.6, 1]$. All null models underestimate the actual P_d, with CM1 being the closest match, followed by CM2, MinDensity and MaxEntropy. I note that, for relatively smaller values of the initial

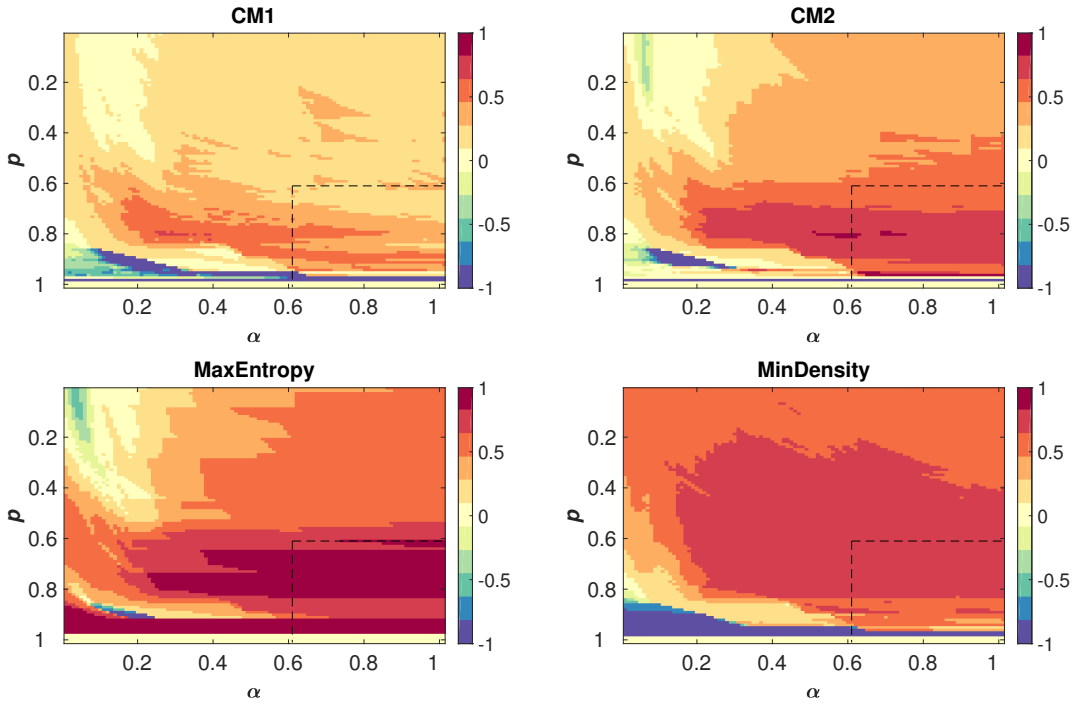
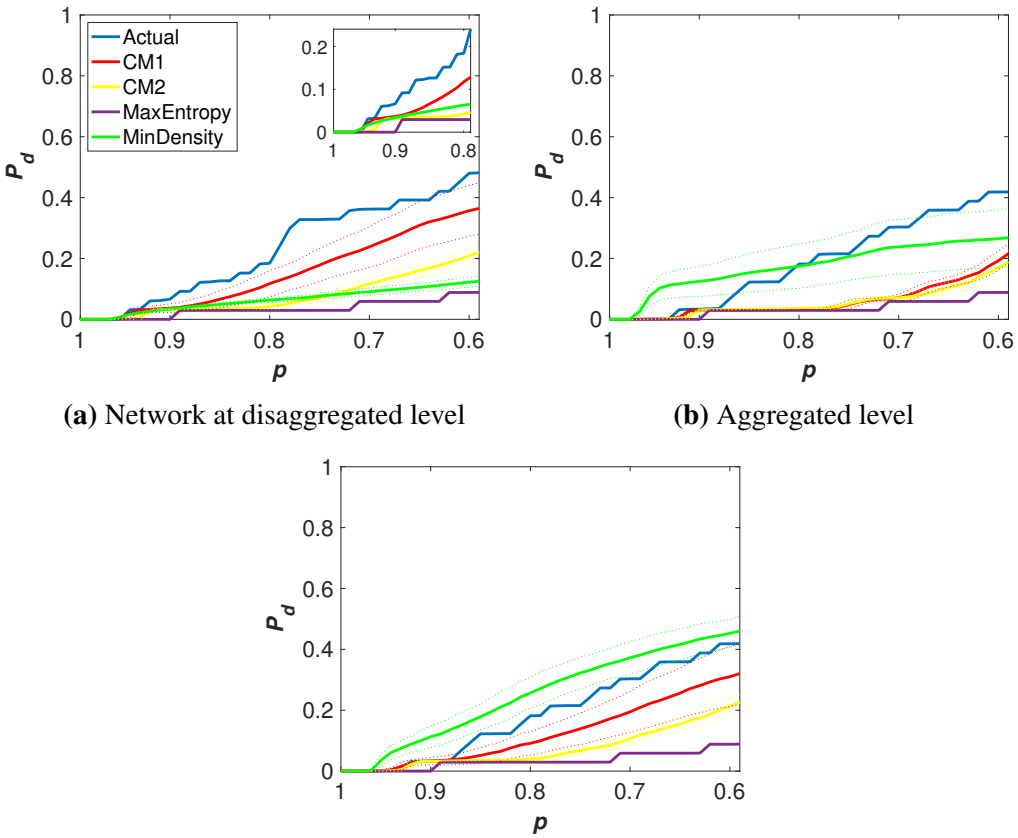


Figure 4.5: Relative difference of the probability of default between actual network and the null models (D_r) at the **disaggregated level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm colour corresponds to an underestimation of the actual network, while cold colour indicates an overestimation. My main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.

shock (inset in Figure 4.6, Panel (a)), MinDensity instead produces higher values of P_d than CM2. This result is driven by the fact that MinDensity produces very sparse networks and allocates very few assets to each bank (high concentration levels). Banks are therefore vulnerable to idiosyncratic shocks in this case, and some banks can default due to the initial shock. However, given that the portfolio overlap between banks is very low in the MinDensity model, shocks cannot spread easily through the system and the number of banks defaulting through fire sale cascades is rather small. The opposite is true for CM2: when the initial shock is large enough to cause banks to default, the shock can propagate to other banks. Therefore, for larger initial shocks, CM2 produces higher values of P_d than MinDensity. Lastly, Panel (a) of Figure 4.7 shows the P_d over time for specific parameters ($\alpha = 0.7$ and $p = 0.6$). As for the 2010 data, the actual network tends to be the most risky one.



(a) Network at disaggregated level

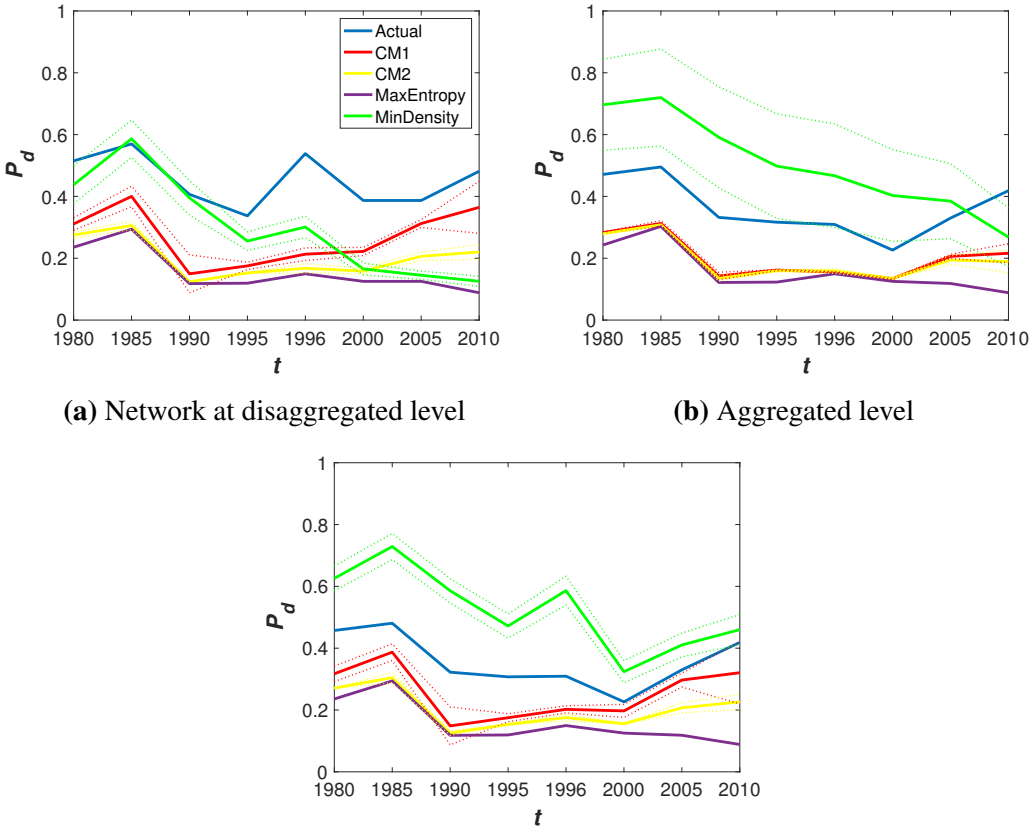
(b) Aggregated level

(c) Intermediate level

Figure 4.6: P_d for initial shock $p \in [0.6, 1]$ and $\alpha = 0.7$. Data for year 2010. Dotted line indicates the value within one standard deviation. Inset: P_d for $p \in [0.8, 1]$ and $\alpha = 0.7$ for network at disaggregated level.

4.3.2.4 Aggregated level

The results for the aggregated networks are shown in Figure 4.8 and in Panel (b) of Figures 4.6 and 4.7, respectively. With the exception of MinDensity, the actual network tends to be the riskiest (at least for the 2010 data) and all other reconstruction methods tend to underestimate the actual P_d for most values of p and α. As for MinDensity, it overestimates (underestimates) the actual P_d for relatively small (large) initial shocks. However, it should be noted that for the aggregated networks, MinDensity produces the riskiest networks in terms of P_d in most years (see also Figure 4.7, Panel (B)). The intuition for this finding is similar to my explanations for the disaggregated networks, with the important difference that MinDensity tends to produce denser networks at the aggregate level. Therefore, the initial shock can



(a) Network at disaggregated level

(b) Aggregated level

(c) Intermediate level

Figure 4.7: P_d over time for illiquid market impact $\alpha = 0.7$ and initial shock $p = 0.6$ for data of different years. Dotted line indicates the value within one standard deviation.

lead to bank defaults which then spread the shock trough the system.

The results also suggest that the other reconstruction models (CM and MaxEntropy) tend to produce values of P_d closer to each other. This suggests that data aggregation may reduce differences among P_d's of different null models. This is mainly because aggregating data will result in a smaller number of nodes (assets) in the network, so reconstruction models have fewer links to allocate. This reduces differences between the reconstructed networks.

4.3.2.5 Intermediate level

For the intermediate aggregation level, the black dashed line in Figure 4.9, and Panel (c) of Figure 4.6 show that MinDensity heavily overestimates the actual P_d. Hence, MinDensity yields the riskiest networks at this aggregation level. The Panel (c) of Figure 4.7 also shows that these results are consistent over time. The reasoning for

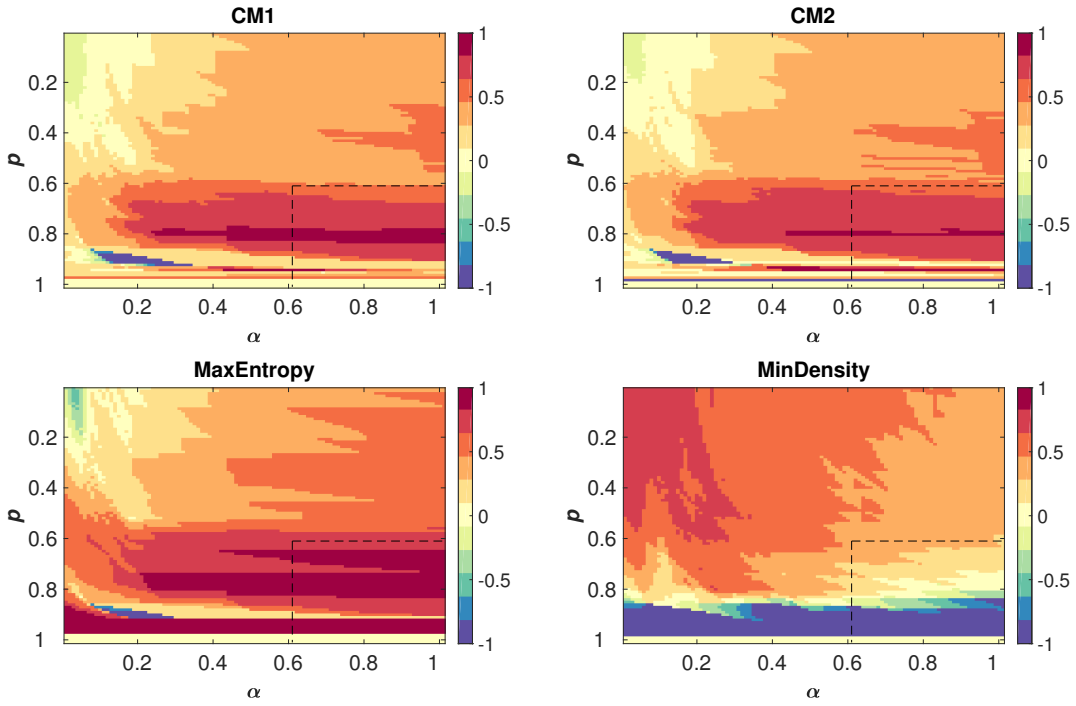


Figure 4.8: Relative difference of the probability of default between actual network and the null models (D_r) at the **aggregated level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm colour corresponds to an underestimation of the actual network, while cold colour indicates an overestimation. My main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.

this finding is again similar to what I saw at the other aggregation levels. The main difference here is that the aggregation takes place after the MinDensity bank-firm networks have been generated, which increases connectivity between banks and thus allows shock to propagate more easily.

4.3.2.6 Summary and discussion: systemic risk analysis

As for the network reconstruction part, Table 4.7 summarizes the results from the systemic risk analysis. I rank the different methods, along with the actual networks, based on the average P_d (standard deviations in parentheses) for the restricted parameter ranges ($p \in [0.6, 1]$ and $\alpha \in [0.6, 1]$).[18]

[18] I also compute the rank for all possible combinations of parameters, including those within and outside restricted range, in Table A.1. I find that the main results are qualitatively similar to those in the main text. I also formally test whether the difference between each network P_d is significant. Specifically, I run a two-sided Wilcoxon signed rank test on each pair of the actual network and the

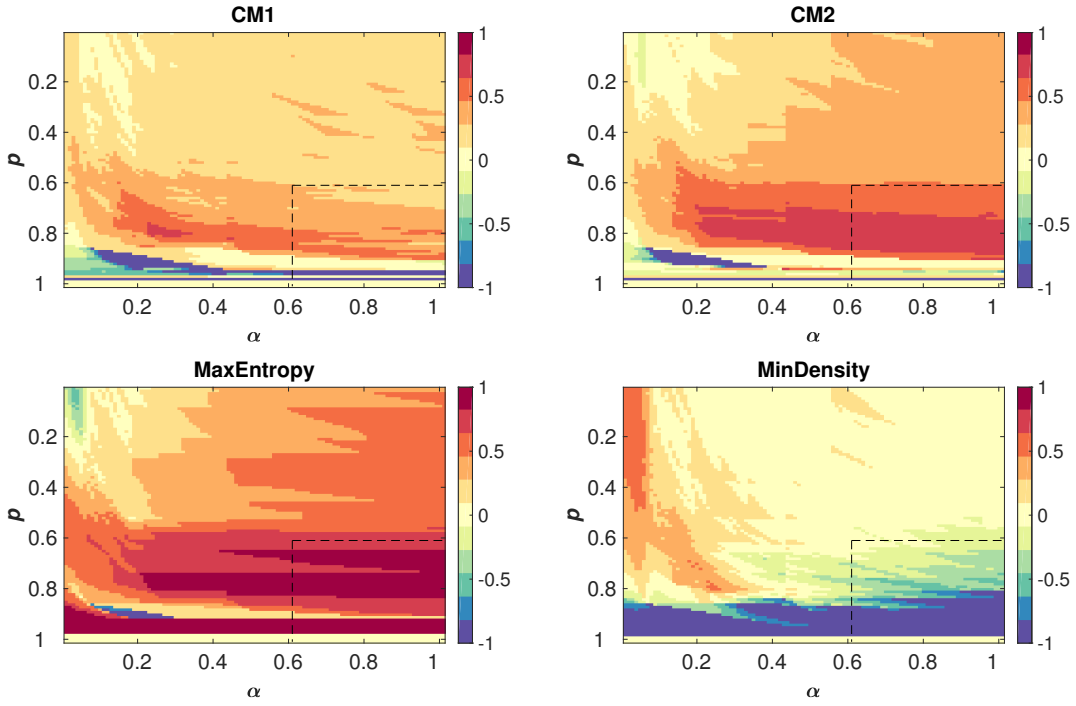


Figure 4.9: Relative difference of the probability of default between actual network and the null models (D_r) at the **intermediate level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm colour corresponds to an underestimation of the actual network, while cold colour indicates an overestimation. My main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.

Rank	Disaggregated		Aggregated		Intermediate	
	Null model	$\overline{P_d}$	**Null model**	$\overline{P_d}$	**Null model**	$\overline{P_d}$
1	Actual	0.236	Actual	0.195	MinDensity	0.275
		(0.164)		(0.148)		(0.159)
2	CM1	0.163	MinDensity	0.193	Actual	0.195
		(0.128)		(0.085)		(0.148)
3	CM2	0.079	CM1	0.061	CM1	0.133
		(0.067)		(0.060)		(0.109)
4	MinDensity	0.069	CM2	0.057	CM2	0.076
		(0.041)		(0.051)		(0.067)
5	MaxEntropy	0.035	MaxEntropy	0.035	MaxEntropy	0.034
		(0.026)		(0.027)		(0.027)

Table 4.7: Rank of the actual networks and the corresponding null models at different aggregation levels for the 2010 data. Rank 1 corresponds to the most risky network. $\overline{P_d}$ denotes the average. I also show the standard deviation of P_d in brackets, which is calculated using the P_d across the restricted parameter range ($p \in \{0.60, 0.61, 0.62, \ldots, 1\}$ and $\alpha \in \{0.60, 0.61, 0.62, \ldots, 1\}$).

First, I find that the actual network tends to display the highest levels of systemic risk in many instances, at least for the disaggregated networks. This is remarkable, given that some of the reconstruction methods generate very different network architectures; for example, MaxEntropy (MinDensity) yields a maximally (minimally) connected credit network. This finding also suggests that even the null models that preserve the degree distribution, like CM1 and CM2, fail to accurately reproduce the actual P_d.[19] However, my result contrasts Anand et al. (2015) which indicates that MinDensity yields an upper bound of the actual risk. Here I find that MinDensity in many instances underestimates the actual P_d, in particular for the disaggregated networks, but overestimates systemic risk at the intermediate aggregation level.

Second, with regard to the performance of each null model, I find that CM1, followed by CM2 and MaxEntropy, has the closest behaviour to the actual network overall, while MinDensity shows an inconsistent performance across different aggregation levels. Given that the different null models require different inputs, I propose CM2 as the most appealing model as it requires less information than CM1.[20]

Lastly, the choice of aggregation level of financial networks matters for stress testing. Certain models can change their behaviour at different aggregation levels, most notably MinDensity. On the other hand, configuration models generally behave rather well at all aggregation levels. However, the ranking of each null model in term of reconstructing the topological features of the actual network is not necessarily consistent with that of reproducing actual systemic risk level (see the comparison between Figure 4.2 and Table 4.7). Future research should explore which network characteristic is most important for reproducing the actual systemic risk levels.

---

null model (see Tables A.1 and Table A.2 in the Appendices for the test results).

[19] This finding is related to previous studies on interbank networks (Mistrulli (2011) and Anand et al. (2015)) which suggest that MaxEntropy underestimates the actual risk.

[20] Wilcoxon tests (see Appendix) indicate that the P_d results from CM1 and CM2 are not significantly different from each other (as opposed to the results from the other reconstruction methods).

4.4 Policy Exercise

My findings suggest that, with respect to the null models I considered, the actual network displays the highest level of P_d in many instances (at least for the 2010 data). This implies that it is possible to make the network more stable by changing its structure. With this in mind, I now explore different policies in order to increase the robustness of the actual credit network.

4.4.1 Policies

To this end, I use a similar approach as Greenwood et al. (2015) and explore three different sets of policies (see Table 4.8 for an overview):

1. merging banks with certain characteristics;

2. breaking up banks with certain characteristics;

3. imposing a leverage cap.

First, I explore the effect of merging banks. In this context, I consider four different scenarios in which I merge a group of large or small banks that are chosen on the basis of their size or leverage. I sort the banks according to their total assets (or their leverage ratios), and merge the top and bottom 15% of them into a single bank.

Second, I study the effect of breaking up banks. Specifically, I split a large bank into two smaller banks. Moreover I assume that one of the smaller banks connects only to the group of relatively connected firms, while the other bank connects only to a group of relatively unconnected firms. (Here I use the number of bank relationships per firm as a proxy of firms' connectedness). Additionally, I assume that the leverage ratio of both banks is identical to the leverage of the original bank.

Third, I explore the effect of a leverage cap, i.e. I limit the maximum ratio between debt to equity of a bank. In this case, I assume that banks that breach the limit need to raise new equity to satisfy the cap (without changing the size of the credit portfolio).

Policy choice	Observable outcome	
1 - Bank merger	Number of banks merged	Total assets of a new merged bank
A) Top 15% (total assets)	17	¥23.58 Tr
B) Top 15% (leverage)	17	¥2.99 Tr
C) Bottom 15% (total assets)	17	¥0.03 Tr
D) Bottom 15% (leverage)	17	¥3.18 Tr
2 - Bank break-up	Number of banks split	Total assets of impacted banks
A) Split each of the top 15% banks (total assets) into one that connects to top 15% industries (connectedness), while the other connects to the bottom 85% industries (connectedness)	17	¥23.58 Tr
B) Split each of the top 15% banks (leverage) into one that connects to top 15% industries (connectedness) and the other connected to the bottom 85% industries (connectedness)	17	¥2.99 Tr
3 - Leverage cap	Equity issue	Number of banks capped
A) max debt/equity = 15	¥354.6 Bn	107
B) max debt/equity = 20	¥79.6 Bn	64
C) max debt/equity = 25	¥34.4 Bn	31
D) max debt/equity = 30	¥18.5 Bn	11

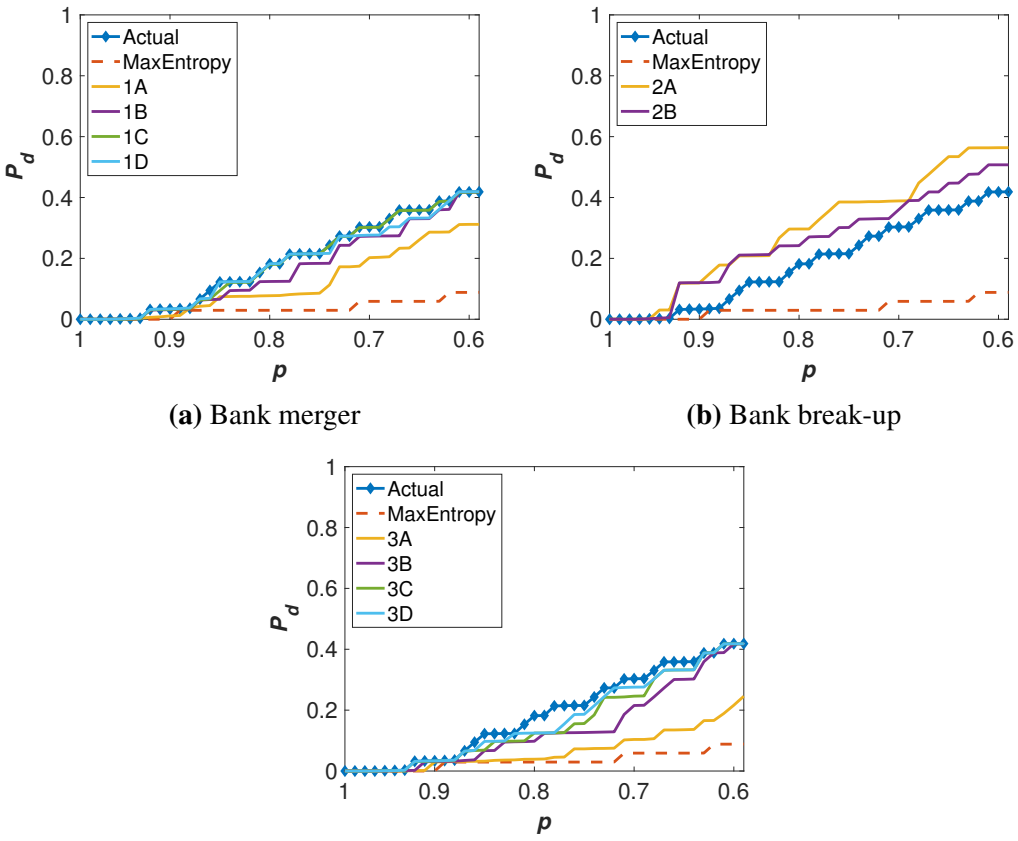
Table 4.8: Different policy exercises applied to the actual network in 2010. Tr and Bn stand for trillion and billion (in ¥).

4.4.2 Results

I apply each policy separately to the actual network and then conduct the systemic risk analysis on these modified networks. For this exercise, I explore the aggregated network in 2010, with $\alpha = 0.7$. I compare P_d of the modified networks to that of the actual network, and to MaxEntropy (the least risky network in this case).

Figure 4.10 shows the results for the three sets of policies. First, I find that merging the largest banks based on their total assets (1A) decreases P_d.[21] My specific merging procedure yields a large but moderately leveraged bank. I illustrate this in Figure 4.11, where the merged bank (red coloured bar) ends up holding 84% assets in the system and with a leverage ratio of 18. Moreover, by holding a majority of the assets in the system, it becomes less vulnerable to other banks' asset liquida-

[21] This is different to the results of Greenwood et al. (2015) where the merger may lead to an even more leveraged bank.



(a) Bank merger **(b)** Bank break-up

(c) Bank leverage cap

Figure 4.10: Effect of different policy exercises on P_d, relative to the actual network. MaxEntropy serves as the lower bound as it is the least risky network in this case. Here I use the data for 2010 with a market impact parameter of $\alpha = 0.7$.

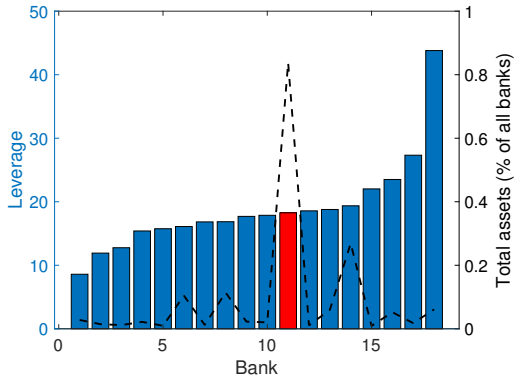


Figure 4.11: Leverage of the top 15% banks (total assets) for the 2010 data. There are 17 (out of 116) banks that belong to this category. The red bar refers to the leverage of a merged bank (resulted from merging the largest 15% banks), while the dashed black line refers to the total assets of the corresponding banks. Overall, the merger results in a very large but moderately leveraged bank.

tions.[22] Figure 4.10 shows that other merging procedures (1B, 1C and 1D) do not lower P_d as effectively. These results are mainly driven by the relatively insignificant total assets of the merged banks obtained from procedures 1B, 1C and 1D as shown in Table 4.8.[23] Moreover, I note that 1B is better than 1D to reduce the actual P_d. The intuition is similar to that of procedure with 1A: merging highly leveraged banks yields a moderately leveraged bank that is more stable during distress.

Figure 4.10 shows that breaking-up banks (2A and 2B) can increase systemic risk. Intuitively, one would expect that this policy should reduce the possibility of shock propagation since I break-up a given bank into two smaller banks that connect to different sets of firms/industries. However, this policy also leads to relatively concentrated banks that are more vulnerable to idiosyncratic shocks. As I only split the top 15% banks and keep the other 75% as they are, banks are still sufficiently interconnected to propagate the shock in the network.

Lastly, Figure 4.10 shows that a leverage cap can lead to substantially more stable networks, with tighter constraints yielding lower values of P_d. However, the results show that for modest leverage caps (such as scenario 3D) P_d remains largely unaffected. Hence, a substantial part of the observed vulnerability of the system is driven by banks' size and their portfolio overlap.

Overall, I find that neither of the three different policy exercises is able to bring down P_d to the values of the least risky network for this particular network (Max-Entropy). I find that merging banks and introducing a leverage cap may improve the robustness of the system, while splitting banks does not. These results can be dependent on the specific choice of the initial shock scenario.

4.5 Conclusions

There is widespread interest in finding accurate reconstruction methods for financial networks from partial information. In this chapter, I focus on reconstructing and

[22] Contrary to Greenwood et al. (2015), I assume that banks only sell assets when they default. Under alternative assumptions (such as leverage targeting), the results may differ from what is reported here.

[23] The leverage of banks in my datasets does not correlate to their size where highly and lowly leveraged banks might consist of both large and small banks.

stress testing bipartite credit networks using detailed micro-data on bank-firm credit interactions in Japan for the period 1980 - 2010.

I find that there is no single ”best” network reconstruction method - it depends on the assumed criterion of interest. This is also true when I look at each method's ability to reproduce observed levels of systemic risk. In fact, in many instances the actual credit networks display the highest levels of systemic risk, at least for the most disaggregated data. Hence, many reconstruction methods tend to underestimate systemic risk. Lastly, I find that the network aggregation level affects the individual performance of the different reconstruction methods.

My findings suggest several interesting paths for future research. First and foremost, it is important to perform similar analyses on other datasets. Secondly, another important follow-up question is whether there are other reconstruction methods that are able to replicate the actual systemic risk levels more closely. In this chapter, I only include a small number of popular reconstruction methods, but other methods may work better. Lastly, different stress tests can lead to different results. I therefore aim to generalize the modelling framework proposed here and test the robustness of the results in future research.

Chapter 5

On Dynamics of Contagion

5.1 Introduction

In this chapter, I consider stress testing models of indirect contagion in a banking network of common asset holdings. For these models, the literature proposed two alternative classes of liquidation dynamics, namely (i) the model of Huang et al. (2013) and Caccioli et al. (2014), and (ii) the model of Greenwood et al. (2015). In the former, banks are assumed to sell their assets only after they have defaulted (*threshold dynamics*). In the latter, banks are assumed to sell their assets whenever their leverage ratio is off-target (*leverage targeting dynamics*). In order to allow for the fact that the actual liquidation behaviour of banks might lie somewhere in-between these extremes, I propose a fire sale model that interpolates between them. The model contains a one-parameter (γ) family of non-linear functions, which determines the volume of assets that a bank liquidates in response to a loss. Intuitively, modest (large) values of γ can be interpreted as the tendency of banks to follow leverage targeting (threshold) dynamics.

I then use the model to predict actual bank defaults for a range of liquidation dynamics, i.e., different values of γ. My goal is to identify the value(s) of γ that perform(s) best in terms of predictive accuracy with regard to actual bank defaults and non-defaults. Following Huang et al. (2013), I use U.S. commercial bank balance sheet data for the last quarter of 2007, and I apply a shock that is meant to mimic the onset of the subprime crisis. I then assess whether the different models manage to

accurately predict the actual defaults that occurred during the years 2008-10 based on the list of bank failures published by the Federal Deposit Insurance Corporation (FDIC).

My main findings are as follows: the performance of the stress testing model strongly depends on the type of initial shock being imposed. On the one hand, systematic shocks tend to yield relatively poor results in terms of predictive power. On the other hand, idiosyncratic shocks can yield much better results, but strongly depend on which asset class is being shocked initially. My approach allows identifying those asset classes that appear most relevant. Based on this identification, I find that the stress testing model displays better performance than a random benchmark, irrespective of the assumed liquidation dynamics. The model is also superior, in most instances, to a standard logistic regression model with leverage and total assets as the sole explanatory variables, which does not account for the network of common asset holdings. Furthermore, I find that the optimal liquidation dynamics depend on the other model parameters, namely the size of the initial shock and the level of market liquidity. I also discuss the fundamental differences between network models and statistical/econometric models in general, and argue that the former are more appealing to the application of systemic stress tests. Lastly, I show that allowing for asset class-specific market impact parameters can improve the model performance, while accounting for multiple rounds of asset liquidation can affect the performance of the model. In particular, I show that considering only the first round of asset liquidations appears most accurate for a model with small γ (banks target their leverage), while accounting for multiple rounds of asset liquidations provides better results for larger γ (banks only liquidate in case of default).

The remainder of this chapter is structured as follows: in section 5.2, I introduce a generalised stress testing model that captures a wide range of behavioural assumptions with regard to bank's liquidation dynamics under stress. In this section, I also provides details on the datasets and experimental setup being used in the research. Section section 5.3 contains the model application and section 5.4 concludes.

5.2 Methodology

5.2.1 Model

In chapter 4, I mainly assumed that banks follow the threshold dynamics, that is banks would only liquidate their assets after default. As I previously discussed in chapter 3, an alternative to this approach would be the leverage targeting dynamics, where banks actively maintain their leverage ratios. In this respect, one might hypothesize that the behaviour of banks might actually lie in-between these two extreme dynamics. To interpolate between the leverage targeting and threshold dynamics, I propose the function $G_i^1(\phi)$ in Equation 3.10 - 3.12 to be defined as:

$$G_i^1(\phi) = \min\left[e^{\gamma\left(\phi_i^1 - \frac{1}{\lambda_i^0}\right)}, 1\right], \tag{5.1}$$

where $\gamma \in (0, \infty)$ is a free parameter that is related to a bank's propensity to follow threshold liquidation dynamics. For example, by setting $\gamma = 0$, I recover the leverage targeting model, in which the amount of assets that a bank liquidates is linear in its losses. For $\gamma > 0$, the response of the bank is non-linear, and it has a convex shape. This means that the bank will increase its rate of liquidation as losses increase. Eventually, in the limit $\gamma \to \infty$, I recover the threshold model. This is shown in Figure 5.1, where I present a comparison between different values of γ in term of liquidation volume as a function of absolute return.

5.2.2 Data

Following Huang et al. (2013), my data come from two sources: first, I take U.S. commercial banks' balance sheet data from Wharton Research Data Services[1] for 2007-Q4. The data set contains the balance sheet of N =7,783 active U.S. commercial banks, with holdings broken up into $M = 14$ broad asset classes (see Table 5.1 for a list of asset classes). This gives us the empirical equivalent of matrix W, which is the bipartite financial network with dimension 7,783 banks $\times$ 14 assets. I also have information on the total assets, total liabilities, and equity of each bank.

[1] https://wrds-web.wharton.upenn.edu/wrds/

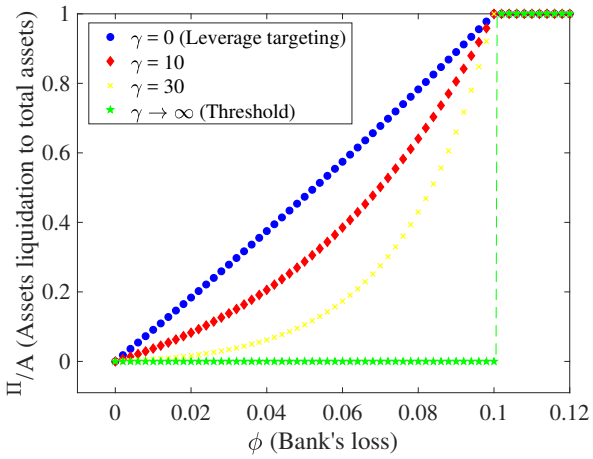


Figure 5.1: The volume of assets liquidated (relative to total assets) as a function of the loss in value of assets for a portfolio with an initial leverage of 10, for different values of γ in the model.

Note that most of the asset classes considered here are loans, and therefore are clearly illiquid. However, I should stress that many of these can indeed be traded on secondary markets (see, e.g., Drucker and Puri (2009)).[2]

The second data set is a list of bank failures from the Federal Deposit Insurance Corporation (FBL-FDIC)[3] for the period 1/1/2008 - 7/1/2011. During this period a total number of 370 banks failed; for 306 of these I have the corresponding balance sheet data from the first database. Hence, roughly 4% of the banks in my sample defaulted during the financial crisis period.

In the following, I apply my stress testing model to the bipartite network from the first data set. I then compare the list of banks that the model predicts to default and the actual list of bank failures. I aim to identify the best performing model that correctly classifies both the 4% of banks that defaulted **and** the 96% of banks that did not default. In other words, I wish to find the value(s) of γ that replicate(s) as closely as possible the behaviour of U.S. banks during the crisis. Fundamentally, my work is therefore related to the growing literature on reverse stress testing (e.g., Grigat and Caccioli (2017)), which aims to identify scenarios that would lead to a certain stress testing outcome. Here the outcome is the default/non-default of

[2]For example, Keys et al. (2010) pointed out that the market of securitized mortgage loans reached $3.6 trillion prior to the crisis.

[3]https://www.fdic.gov/bank/individual/failed/banklist.html

Index	Asset class	Total amount (billion USD)	Relative amount (% tot. assets)
1	Loans for construction and land development	774.67	6.10
2	Loans secured by farmland	247.54	1.95
3	Loans secured by 1-4 family residential properties	3058.64	24.09
4	Loans secured by multi-family (> 5) residential properties	185.11	1.46
5	Loans secured by non-farm non-residential properties	1234.33	9.72
6	Agricultural loans	224.31	1.77
7	Commercial and industrial loans	1605.34	12.64
8	Loans to individuals	1351.53	10.64
9	All other loans (exclude consumer loans)	282.80	2.23
10	Obligations (other than securities and leases) of states and political subdivision in the U.S.	98.68	0.78
11	Held-to-maturity securities	772.17	6.08
12	Available-for-sale securities, total	2221.73	17.50
13	Premises and fixed assets including capitalized lease	145.26	1.14
14	Cash	496.93	3.91
	Total assets	**12,699.04**	**100**

Table 5.1: Different asset classes used in this chapter. There are 14 asset classes in total: except for cash, the other 13 asset classes are less than perfectly liquid (asset liquidations generate market impact). The third column is the total amount of each asset class in billion USD, and the fourth column is the corresponding percentage share of total assets.

individual banks.

5.2.3 Experimental Setup

Let us spell out some assumptions and introduce some definitions that I use in this study. I start the cascading process by reducing the value of an asset $j \in \{1,...,M-1\}$[4] to a fraction p of its original value ($p \in \{0, 0.01, 0.02, ..., 0.99, 1\}$. In this sense, a large (small) value of p corresponds to small (large) initial shock. With regard to the market impact, I consider the linear function where I iterate over different values of α ($\alpha \in \{0, 0.01, 0.02, ..., 0.99, 1\}$) and, initially, I assume that this value is homogeneous across assets[5] (except for cash, for which I always consider $\alpha = 0$). I will relax the homogeneity assumption below.

[4]There are M assets in the network, including cash. I would only shock the value of the non-cash asset.

[5]Many studies make a simplifying homogeneity assumption with regard to the market impact function (e.g. Greenwood et al. (2015)). Put simply, this means that all assets are assumed to be equally liquid.

In the following, I will show results for all possible combinations of p and α, but I mainly focus on the results for a restricted range of parameters. Since the majority of asset classes considered here are relatively illiquid, I follow the approach that I previously used in chapter 4 and focus on the upper range of the market impact parameter ($\alpha \in [0.6, 1]$).[6] Moreover, in line with previous studies on indirect contagion (Cont and Schaanning (2017), Greenwood et al. (2015)), I consider relatively small initial shocks ($p \in [0.6, 1]$). This makes intuitive sense, given that network models of systemic risk are based on the idea that relatively small shocks can amplify through the network and thus potentially have large effects overall.

Concerning the number of liquidation rounds, I mainly focus on the first iteration round, which is in line with the approach of Greenwood et al. (2015) and Fricke and Fricke (2020). More specifically, the first three steps of the contagion algorithm described at the beginning of section 3.3 are only performed once, and the optional fourth step is left out. This approach differs from the analysis of Huang et al. (2013), where the contagion algorithm is iterated until convergence. I later compare the model accuracy when allowing for multiple rounds of asset liquidations.

Concerning the propensity of banks to follow threshold dynamics (parameter γ), I consider values of $\gamma \in \{0, 1, ..., 49\} \cup \{\infty\}$. Recall that a value of $\gamma \to \infty$ means that banks exactly follow threshold dynamics, i.e. they only liquidate assets in case of default. On the other hand, lower values of γ indicate a tendency to follow leverage targeting dynamics. For each value of γ from my set of values, I perform a stress test and measure the prediction accuracy of the corresponding model. To this end, I evaluate each model's ability to identify bank failures and non-failures during financial distress correctly. Here I use the standard receiver operating characteristic (ROC) curve (Egan (1975); Swets et al. (2000)) which shows the fraction of correctly identified bank failures (true positive rate/TPR) versus the incorrectly classified failures (false positive rate/FPR). A random classifier would yield a di-

[6] For $\alpha = 0.6$, the asset price drops by 6% when 10% of the asset is liquidated; for $\alpha = 1$, the price drops by 10% when 10% of the asset is liquidated.

agonal line (from bottom left to top right) in the ROC space; see Figure 5.2. In contrast, a perfect model produces points closer to the top left corner in the ROC space.

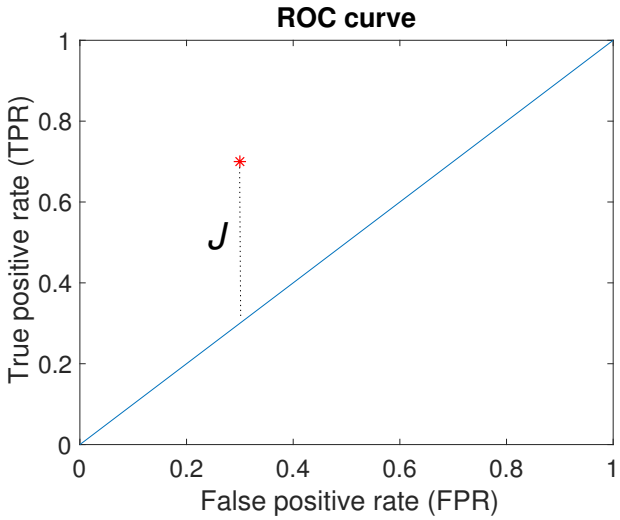


Figure 5.2: Illustration of Youden's J statistic. The red point in the ROC space corresponds to FPR = 0.3 and TPR = 0.7. The value of J of this point is displayed by the dotted line ($J = 0.7 - 0.3 = 0.4$). Larger values indicate a better model performance.

In the following, each value of γ will produce one ROC curve. Each point corresponding to this curve corresponds to one combination of p and α, i.e., the size of the initial shock and asset illiquidity. In order to test the model accuracy for each combination of p and α I use *Youden's J statistic* (Youden (1950)), which is defined as :

$$J_{\gamma,p,\alpha} = \text{TPR}_{\gamma,p,\alpha} - \text{FPR}_{\gamma,p,\alpha}. \tag{5.2}$$

The value of J ranges from 0 to 1. $J = 0$ indicates the performance of a random classifier, while $J = 1$ denotes the performance of a perfect model. Figure 5.2 illustrates the computation of J.

Unless otherwise stated, in my analysis below I always exclude bankruptcies due to the initial shock and focus on failures due to the contagion process only. This is to separate the effect of shock propagation/amplification via the fire sale mechanism from the influence of the initial shock. (In the Appendix, I take a closer look at defaults due to the initial shock.) Moreover, to compare the accuracy of

the network models with some benchmarks, I also plot two other ROC curves in each figure: i) a random classifier, and ii) a standard logistic regression model with bank leverage and total assets (in logs) as explanatory variables.[7] I believe that this benchmark is a reasonable starting point since it does not include any information on the network of common asset holdings. (In Section 5.3.1.4 I will introduce a more sophisticated benchmark model that does include network information.)

I also note that, in the following analysis, the ROC curve of the logistic regression benchmark and that of the network model come from slightly different samples: in the former, I consider failures of all banks in the data set, while in the latter I only consider failures due to the contagion process and exclude those due to the initial shock. My aim here is therefore to look at the comparison between the accuracy of contagion dynamics and a benchmark that uses bank features as predictive variables. I should highlight that if I also considered failures due to the initial shock for the network model, I would obtain more superior ROC curves compared to those that I show here (see Appendix B.3). Therefore, if anything, in comparison with the logistic regressions, the accuracy of the network model is biased downwards.

5.3 Results

5.3.1 Model Performance

Apart from the model-specfic parameters (in my case: p, α, and γ), the researcher needs to decide what kind of shock should be imposed (e.g., idiosyncratic or systematic). In the case of idiosyncratic shocks, one also has to decide which asset class(es) should be shocked. In the following, I will show how to pick the *relevant* asset classes in terms of how the stress testing model performs in the classification exercise. I will then focus on the results for the most relevant asset classes only (exemplary results for less relevant asset classes can be found in the Appendix). As it turns out, the two relevant asset classes that I identify (asset class 1: loans for construction and land development; asset class 5: loans secured by non-farm non-residential properties) made up less than 15% of banks' total assets before the

[7] See Appendix B.1 for details on the regression model.

crisis. Interestingly, these two asset classes are exactly those that were problematic during the global financial crisis (Cole and White (2012); Huang et al. (2013)).

5.3.1.1 Identifying relevant asset classes

The stress scenario to be studied is to be defined by the researcher. One naïve approach would be to run the model multiple times, with different asset classes being hit by the initial shock and then averaging the results over the different shock scenarios. However, as shown in Figure 5.3, this yields very poor classification results. Specifically, I run the stress testing model separately for each asset class (except for cash) and compute the average of the TPR/FPR across all shock scenarios for each parameter combination. The ROC curves in Figure 5.3 are quite close to those of a random classifier (and substantially below those of a logistic regression), suggesting that this approach has very little predictive power.

Shocked asset	*Complete parameter range*					*Restricted parameter range*				
class (j)	Mean	Std	Min	Max	Prob	Mean	Std	Min	Max	Prob
1	0.276	0.212	-0.000	0.593	0.215	0.241	0.224	-0.000	0.592	0.379
2	-0.034	0.040	-0.149	0.199	0.662	-0.007	0.010	-0.066	0.003	0.997
3	0.011	0.033	-0.139	0.140	0.908	0.017	0.019	-0.064	0.119	0.975
4	0.023	0.032	-0.000	0.347	0.960	0.004	0.004	-0.000	0.042	1.000
5	0.120	0.121	-0.000	0.420	0.433	0.137	0.142	-0.000	0.420	0.420
6	-0.042	0.049	-0.132	0.199	0.545	-0.006	0.007	-0.037	0.010	1.000
7	0.043	0.039	-0.000	0.215	0.687	0.049	0.051	-0.000	0.213	0.575
8	-0.020	0.049	-0.197	0.093	0.774	-0.025	0.045	-0.178	0.047	0.781
9	-0.000	0.011	-0.124	0.033	0.982	0.000	0.002	-0.013	0.030	1.000
10	-0.000	0.002	-0.004	0.033	1.000	-0.000	0.000	-0.000	0.003	1.000
11	0.013	0.045	-0.095	0.207	0.689	0.006	0.035	-0.052	0.198	0.899
12	-0.006	0.053	-0.120	0.112	0.852	0.004	0.035	-0.147	0.110	0.895
13	0.000	0.012	-0.002	0.264	0.993	-0.000	0.000	-0.000	0.003	1.000

Table 5.2: Unconditional average of *Youden's J statistic* (across all combinations of γ, p, and α) when imposing an initial shock on each asset class separately.

In order to disentangle this somewhat disappointing finding, I show separate results for the individual asset classes in Table 5.2. In particular, I compute the average value of *Youden's J statistic* (Equation 5.2) along all (p, α, γ) combinations for each asset class. Note that a random classifier corresponds to $J = 0$, while a perfect classifier would correspond to $J = 1$. In order to assess the significance of the reported J statistic, I also perform a simple simulation-based significance test: first, I simulate the ROC curve for a random classifier with the same number of observations as the data. (With a sufficiently large number of observations, the

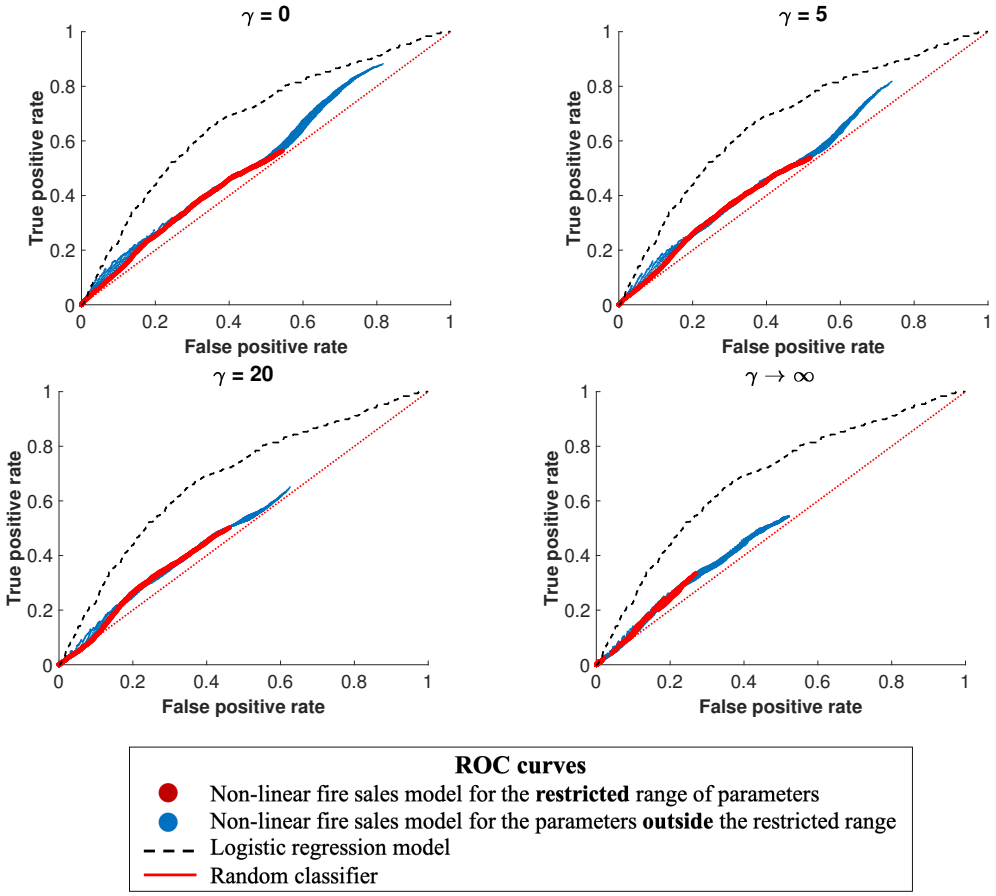


Figure 5.3: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., for the average results of initial shock across all asset classes (except cash). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

expected value of the ROC curve converges to the 45 degree line shown in the Figures.) This gives us the distribution of J for a random classifier. For each asset class j, I can then count how many of the observations fall within the 5% confidence interval of the random classifier. This number is reported as Prob in Table 5.2. Based on this simple significance test, I find that the unconditional J statistic (both for the complete parameter range, and for the restricted parameter range) is not significantly different from zero for any of the asset classes; in fact, for most asset classes, I find that close to 100% of the J statistic fall within the confidence interval of a random classifier. Put differently, imposing an idiosyncratic shock on most of the asset classes yields results that are indistinguishable from those of a random classifier.[8] Table 5.2 also shows that for asset classes 1 and 5 the unconditional results are slightly better: approximately 60% of the J statistic fall outside the 5% confidence bands of the random classifier. While these findings would still not count as statistically significant at reasonable confidence levels, I treat these two asset classes as the most relevant ones in what follows.

Note that my findings on asset classes 1 and 5 are in line with Cole and White (2012), who found that banks with high levels of commercial real estate loans were particularly affected during the recent global financial crisis. Indeed, I find that the classifications are significantly better compared to those of a random classifier for initial shocks on these two asset classes. I will therefore focus on results from these two asset classes in most of what follows (examples of results from less relevant asset classes can be found in the Appendix).

5.3.1.2 Results for the two most relevant asset classes

The different panels in Figure 5.4 show the ROC curves based on $\gamma = 0, 5, 20$ and $\gamma \to \infty$, when imposing an initial shock on asset class 1. Figure 5.5 shows the corresponding results when imposing initial shock on asset 5 instead. Each dot in the ROC curve corresponds to the result of the stress test for a particular combination of p and α. Red dots correspond to the results for the restricted range of parameters,

[8] I also perform standard two-sample Kolmogorov-Smirnov tests to compare the different distributions of the J statistic with the one of the random classifier. Here I reject the null hypothesis for all asset classes.

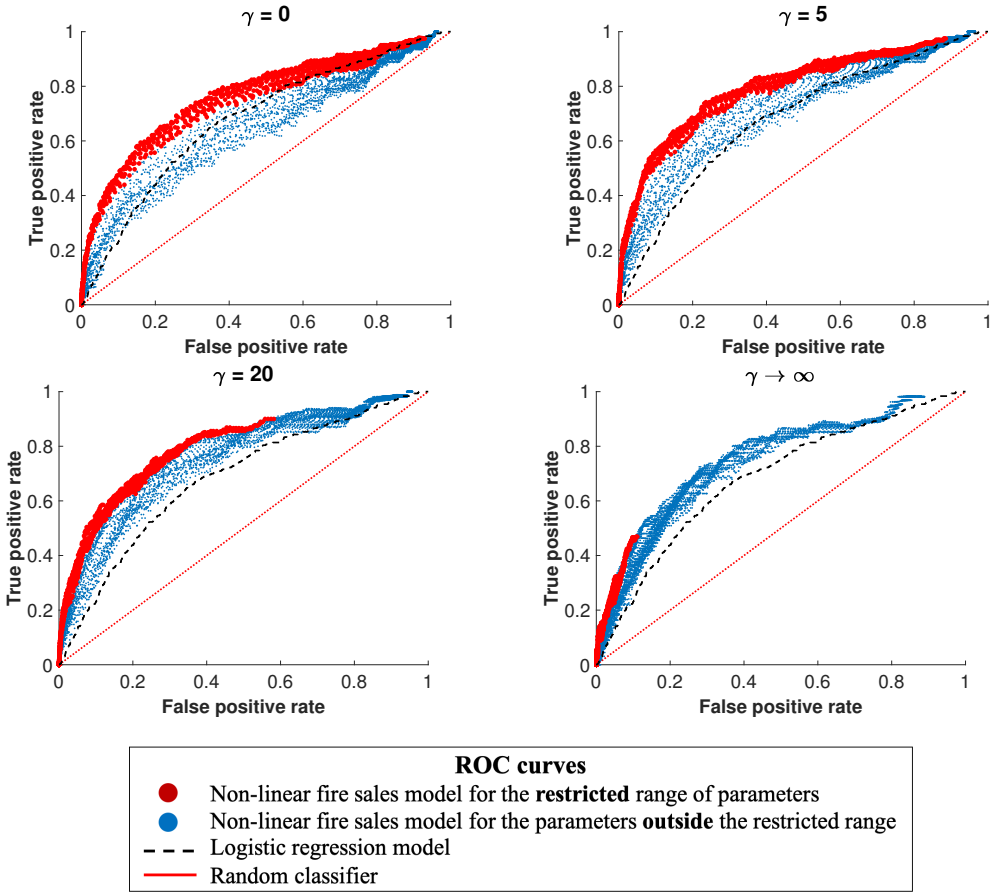


Figure 5.4: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with initial shock on loans for construction and land development (asset class 1). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

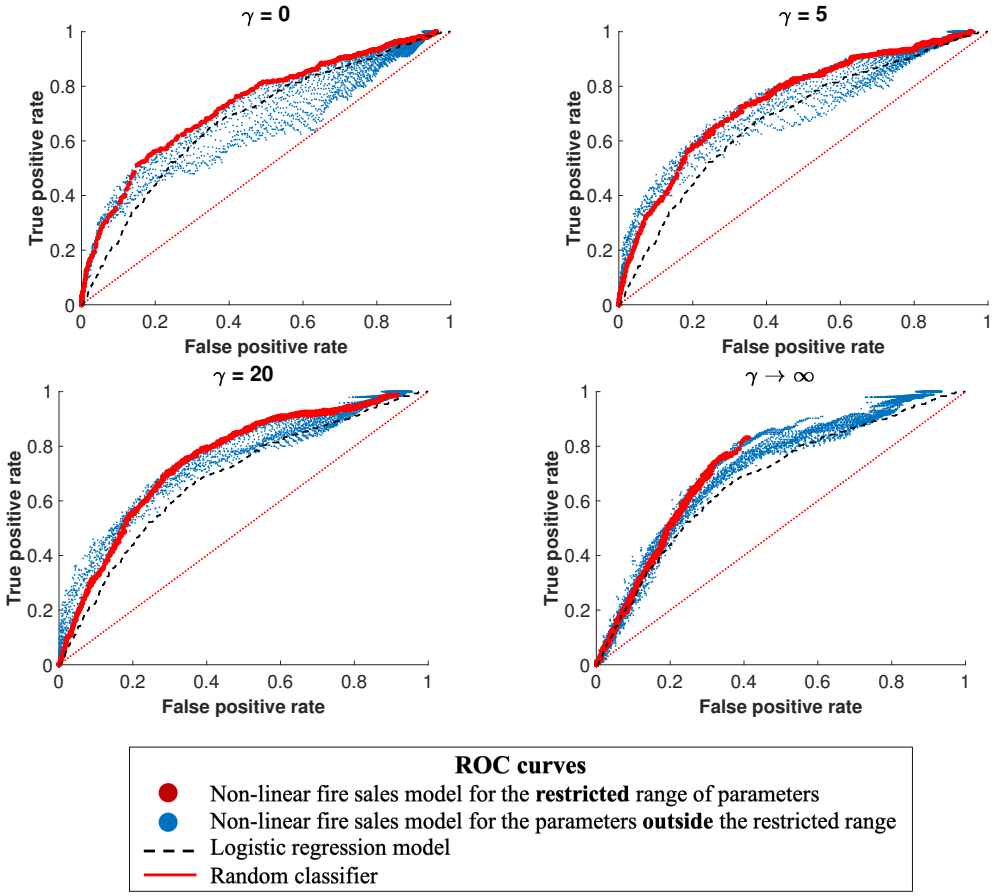


Figure 5.5: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with initial shock on loans secured by non-farm non-residential properties (asset class 5). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

blue dots indicate results outside this range. Overall, I find that the model performs better than a random classifier for all values of γ for these two asset classes. This can be seen from Figure 5.4 and Figure 5.5, where the ROC curves lie above diagonal lines. In most instances, I also find that the model is superior compared to the logistic regression model, which indicates that the network of common asset holdings indeed contains useful information to predict defaults. However, for different values of γ, the same combinations of p and α can end up in different locations of the ROC space. For example, for larger γ (bottom panels in Figure 5.4 and Figure 5.5), the red dots cover only a limited range within the TPR/FPR space. This is due to the fact that values closer to the top right corner of the TPR/FPR space correspond to a larger number of banks that are being predicted to default. Meanwhile, red dots correspond to the results for the restricted range of parameters (low initial shock and high market impact). For higher values of γ, banks are less aggressive in terms of their leverage targeting, and therefore liquidate fewer assets during distress. Hence, when γ is large and the initial shock is small, bank defaults are rare and shocks propagate slowly through the system. For smaller values of γ, this changes dramatically: when banks' propensity to target their leverage ratios is strong, shocks can propagate more easily through the network (see section 5.3.2). As I see from the top panels in Figure 5.4 and Figure 5.5, the red dots now also cover a broader range within the TPR/FPR space. These findings suggest that there may be different regimes in the (p, α) plane, for which different values of γ are best.

5.3.1.3 Model performance in the (p, α) plane

In the following, I now discuss the model performance in the (p, α) plane and check for each combination of (p, α) which value of γ yields the most accurate model. Specifically, I find the value of γ for which $J_{\gamma,p,\alpha}$ is maximized:

$$\gamma^{\text{opt}}(p, \alpha) = \arg\max_{\gamma} J_{\gamma,p,\alpha}. \tag{5.3}$$

As noted previously, here I focus only on the results for the two most relevant asset classes (1 and 5). In the following, I differentiate between results for the complete (p, α) range, and results for the restricted range with relatively small initial shocks and relatively large price impact parameters.

Complete Parameter Range. The values of $\gamma^{\text{opt}}(p, \alpha)$ are shown as a heatmap in Figure 5.6(a) and Figure 5.7(a), for shocks on asset class 1 and 5, respectively. The Figures suggest the existence of three different regions characterized by different values of γ^{opt}: two of these regions (denoted as regions 1 and 3) indicate that the threshold model appears to be best ($\gamma^{\text{opt}}(p, \alpha) \to \infty$).[9] There is also an intermediate region (denoted as region 2) where the best performing value of γ is smaller, such that the leverage targeting model yields more accurate results ($\gamma^{\text{opt}}(p, \alpha) \to 0$). Note that there are also some cases in region 2 where something in-between the threshold model and the leverage targeting approach is best. Overall, it should be clear that the best performing γ varies across different (p, α) combinations.

In order to take a closer look at the accuracy of the best performing model in each regime, Figure 5.6(b) and Figure 5.7(b) show the value of $J_{\gamma^{\text{opt}},p,\alpha}$ in the (p, α) plane. As I previously discussed, a random classifier would correspond to $J = 0$, while a perfect model corresponds to $J = 1$. The results suggest that region 1 is characterized by small values of J, which is incidentally not driven by the model's failure to accurately identify bank defaults (high FPR, but low TPR), but simply because the model does not identify any default (low FPR and low TPR). Since this region is characterized by modest values of (p, α), the exogenous shock would not be adequately amplified to trigger further contagion in the system. On the other hand, regions 2 and 3 display relatively high values of J ($J > 0.4$).

Finally, to get a feeling for the variation in model accuracy for different values of γ, Figure 5.6(c) and Figure 5.7(c) show the differences in performance between the best and the worst model in the (p, α) plane:

$$\delta(p, \alpha) = \max_{\gamma} J_{\gamma,p,\alpha} - \min_{\gamma} J_{\gamma,p,\alpha}, \tag{5.4}$$

[9] I also see a few cases in region 1 where the accuracy of all values of γ is equivalent (shown in white).

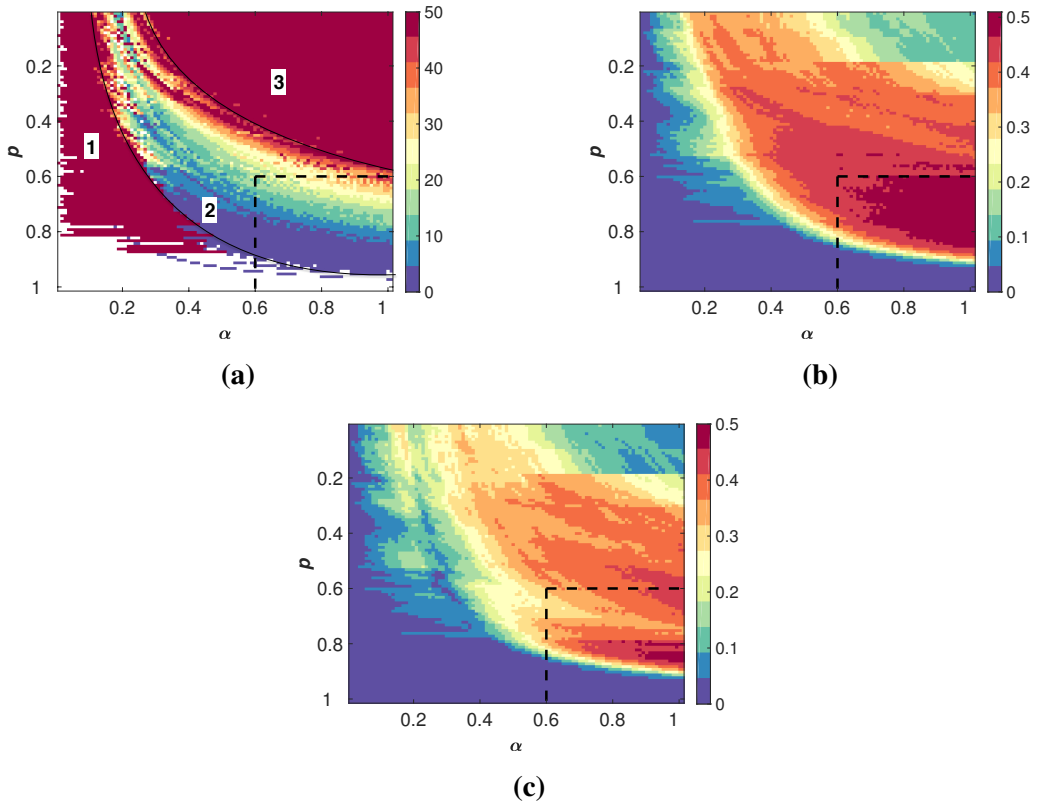


Figure 5.6: Results from imposing an initial shock on asset class 1 (loans for construction and land development).(a) The variation of γ^{opt} (based on Youden's J statistic) in different (p, α) regimes. Warmer (colder) colours refer to larger (smaller) values of γ, and white is used whenever the accuracy of all γ are equivalently similar. I look at the range of $\gamma \in \{0, ..., 49\} \cup \{\infty\}$, and $\gamma = 50$ in this plot corresponds to $\gamma \to \infty$. **(b)** values of J of the corresponding most accurate γ in different regimes of p and α. **(c)** The variation of model accuracy between the values of J of the most accurate and the least accurate γ in different regimes.

Note that δ ranges between 0 (where both models have a similar value of J) and 1 (if one model is a perfect model while the other is a random classifier).

As shown in the figures, I again obtain high values of δ for most (p,α) combinations in regions 2 and 3. Hence, properly tuning γ can substantially improve model accuracy. This also implies that there is no contagion algorithm that is always the best, but rather the best performing value of γ depends on the specific values of p and α. Note that the results are consistent for both of the most relevant asset classes.

A high-level overview of my main findings can be found in Table 5.3. In

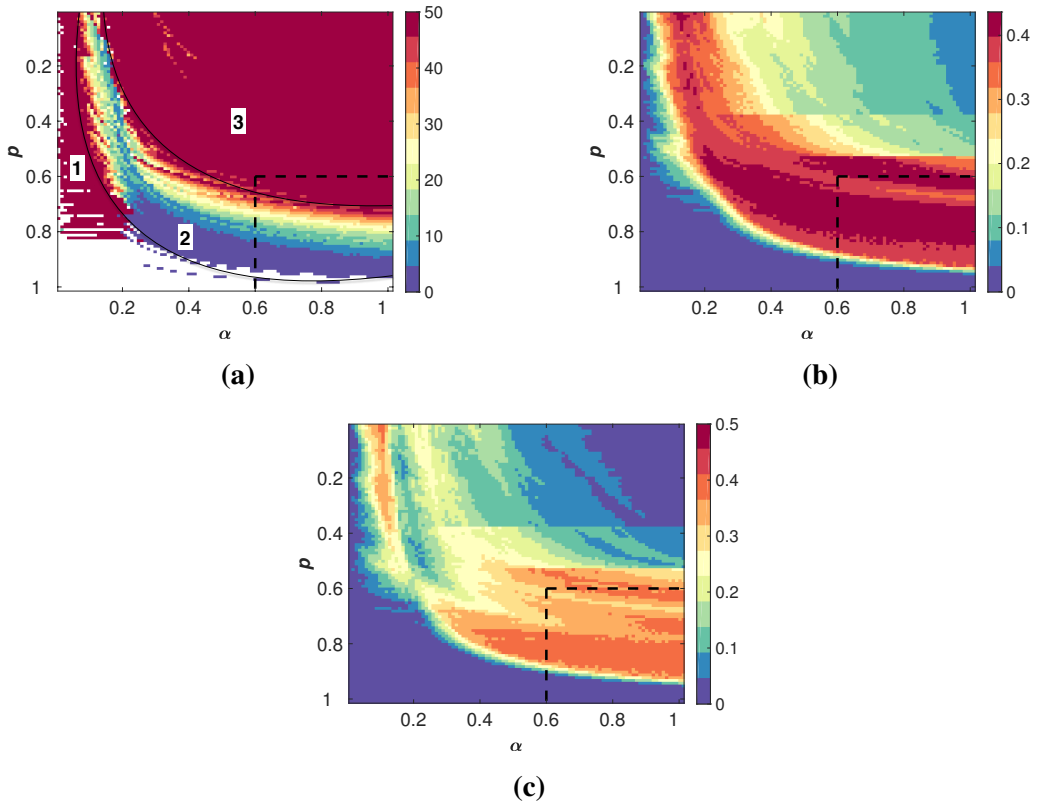


Figure 5.7: Results from imposing an initial shock on asset class 5 (loans secured by non-farm non-residential properties). **(a)** The variation of γ^{opt} (based on Youden's J statistic) in different (p, α) regimes. Warmer (colder) colours refer to larger (smaller) values of γ, and white is used whenever the accuracy of all γ are equivalently similar. I look at the range of $\gamma \in \{0, ..., 49\} \cup \{\infty\}$, and γ = 50 in this plot corresponds to $\gamma \to \infty$. **(b)** values of J of the corresponding most accurate γ in different regimes of p and α. **(c)** The variation of model accuracy between the values of J of the most accurate and the least accurate γ in different regimes.

practice, once the size of exogenous shock and the strength of market impact have been calibrated, one can select the corresponding best performing value of γ that I provided in Figure 5.6 and Figure 5.7.

Restricted Parameter Range. I now take a closer look at the model performance in the regime of small initial shocks ($p \in [0.6, 1]$) and high market impact parameters ($\alpha \in [0.6, 1]$). The results for the restricted range of parameters are highlighted in Figure 5.6 and Figure 5.7 as the area inside the black dashed line square (bottom right corner). Interestingly, Figure 5.6(a) and Figure 5.7(a) show that the best performing liquidation parameter γ in this range lies in-between leverage target-

		Market impact (α)		
		weak $0 \lesssim p \lesssim 0.3$	*moderate* $0.3 \lesssim p \lesssim 0.7$	*strong* $0.7 \lesssim p \lesssim 1$
Initial shock (p)	*large* $0 \lesssim p \lesssim 0.3$	All dynamics mostly the same	Threshold dyamic	Threshold dynamic
	moderate $0.3 \lesssim p \lesssim 0.7$	All dynamics mostly the same	Intermediate region	Threshold dynamic
	small $0.7 \lesssim p \lesssim 1$	All dynamics exactly the same	Leverage targeting	Leverage targeting

Table 5.3: The best performing liquidation dynamics for different combinations of the exogenous shock and the market impact parameter.

ing and the threshold dynamics. Furthermore, I also observe from Table 5.3 that the best performing model in this range switches from the leverage targeting (high price impact, small shocks) to the threshold dynamics (high price impact, moderate shocks).

These results can be explained as follows: with smaller values of γ, banks target their leverage ratios more aggressively. Hence, they will liquidate more assets during distress. When the initial shock is small (e.g. $p = 0.9$), banks only observe a small decline in their total assets and shocks are therefore unlikely to spread. That is, unless banks decide to liquidate a considerable amount of their assets. Accordingly, the shock propagation and the (accurate) default prediction can only be observed for the model with $\gamma = 0$ (leverage targeting).

As the initial shock becomes larger (e.g. $p = 0.7$), the best performing γ shifts to the region that lies in between leverage targeting ($\gamma = 0$) and threshold dynamics ($\gamma \to \infty$). The reason for this is the following: shocks can now spread even if banks only decide to liquidate a moderate amount of their assets, since losses from the initial shock are larger. Accordingly, a more accurate model uses an intermediate value of $0 < \gamma < \infty$. In particular, the leverage targeting model that was best performing for small shocks now overestimates the number of defaults, such that the increase in its false positive rate (the number of banks that it incorrectly predicts to default) is faster than the increase in its true positive rate (the number of banks that it correctly predicts to default). Therefore, models with intermediate γ (e.g.

$\gamma \in [20, 30]$) become more accurate.

5.3.1.4 Comparison with a more sophisticated benchmark

The results suggest that network models have some predictive power, as they predict defaults better than both a random classifier and the simple logistic regression benchmark. In this section, I compare the performance of the model with that of a more sophisticated logistic regression, where I also include banks' relative amount of holdings of each asset class (defined as the portfolio share). To deal with collinearity, I exclude asset classes 10 and 13 in the regression, which account only for less than 2% of the aggregated balance sheet. In other words, I incorporate information on the network of common asset holdings. Interestingly, as shown in Table B.2 in the Appendix, the parameter on log(Leverage) now turns out insignificant. Only the parameters on log(TotalAssets), and on asset classes 1 and 8 are significant. Despite this, the pseudo-R^2 now increases substantially from 0.01 to 0.17.

In Figures 5.8 and 5.9, I plot the performance of this benchmark against that of the network model when I shock assets 1 and 5 respectively. The results indeed suggest that the enhanced logistic regression appears to perform better than the network model. This, however, should not be taken as a defeat for the network model since the enhanced logistic regression model is subject to strong overfitting to the subprime crisis: the model has been trained to discriminate between banks that defaulted and those that did not after the 2008 shock, which mostly affected asset classes 1 and 5 (Cole and White, 2012). For this type of shock, the model performs quite well, but it would not be able to provide any predictions under slightly different scenarios (e.g., different exogenous shocks), which is however important from a macroprudential policy perspective.

More generally, in contrast to statistical/econometric models, network models have not been developed or optimized for prediction purposes. Rather, network models have primarily been used for developing intuition about the dynamics of contagion, its parameter sensitivity, and to perform scenario analysis. This is particularly true for the growing literature on reverse stress testing. The higher flexi-

bility of the network model can be seen in Figure 5.10, where the exogenous shock affects asset 8. The fact that the network model does not perform better than the random benchmark is due to the fact that the model predicts the default of different banks with respect to the case when assets 1 and 5 are shocked, while the logistic regression always predicts exactly the same banks to default in both cases.

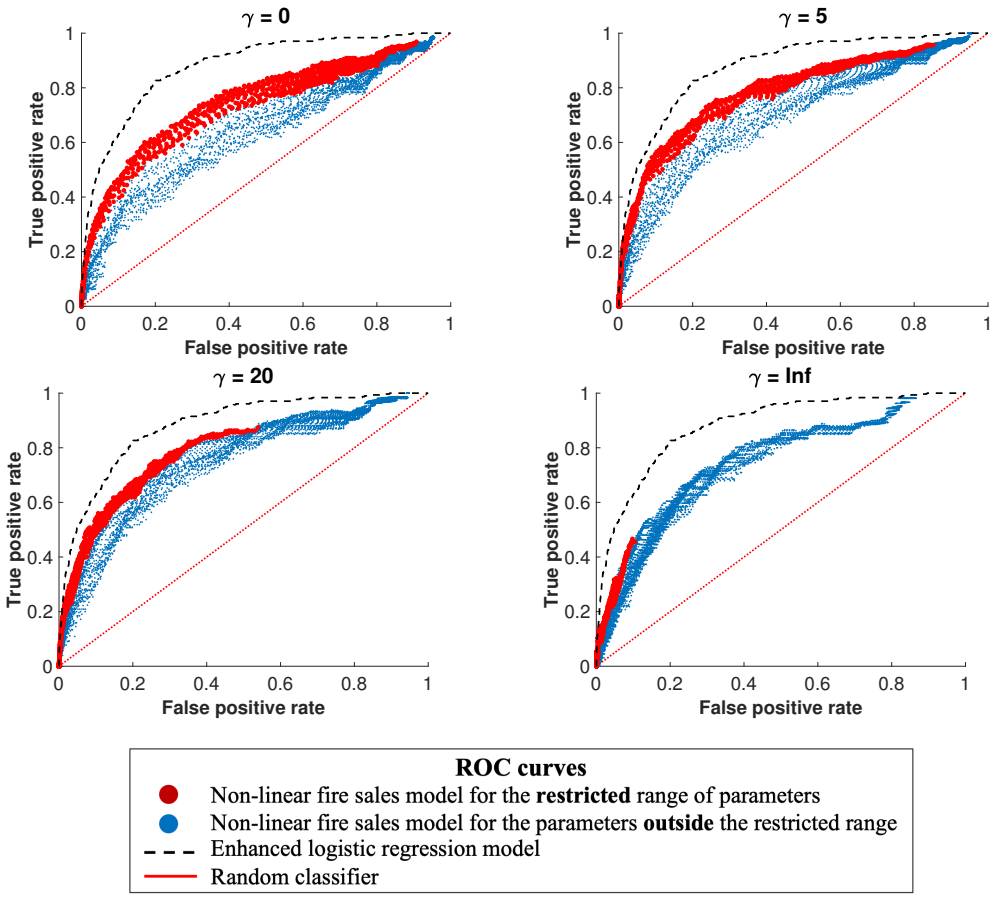


Figure 5.8: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with initial shock on loans for construction and land development (asset class 1). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed lines is the ROC curve of an enhanced logistic regression model with bank leverage and total assets (log-transformed) and investments in all asset classes but 10 and 13 (defined as the portfolio share) as explanatory variables. The red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

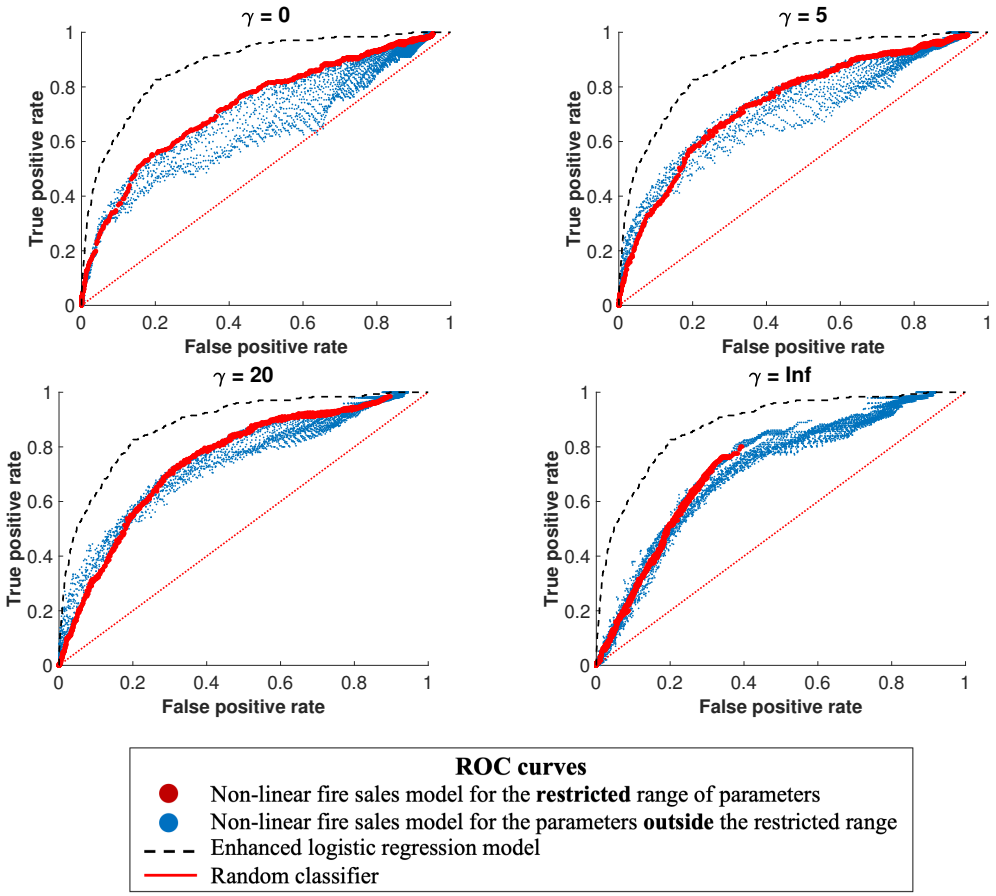


Figure 5.9: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with initial shock on loans for construction and land development (asset class 5). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed lines is the ROC curve of an enhanced logistic regression model with bank leverage and total assets (log-transformed) and investments in all asset classes but 10 and 13 (defined as the portfolio share) as explanatory variables. The red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

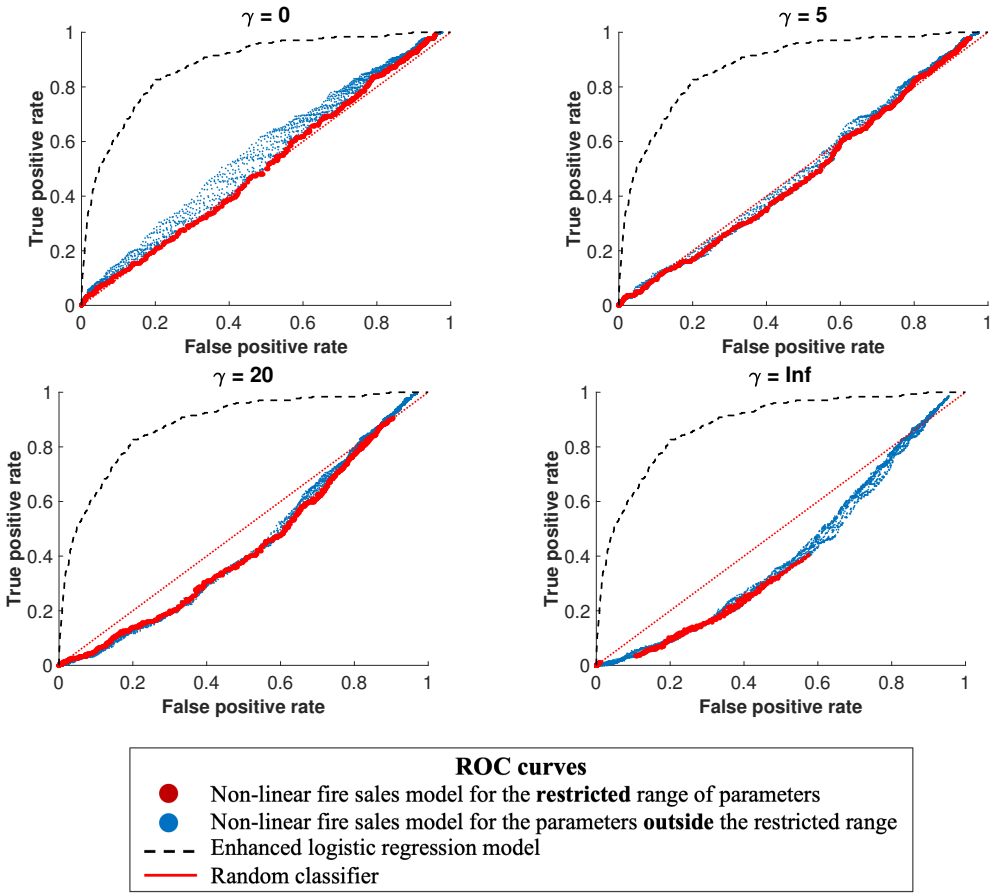


Figure 5.10: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with initial shock on loans for construction and land development (asset class 8). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed lines is the ROC curve of an enhanced logistic regression model with bank leverage and total assets (log-transformed) and investments in all asset classes but 10 and 13 (defined as the portfolio share) as explanatory variables. The red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

5.3.2 Extensions

In the following, I assess to what extent the model performance can be improved by means of two separate model extensions: (1) heterogeneity in the market impact parameter (α_j) across different asset classes; (2) multiple rounds of asset liquidations.

5.3.2.1 Heterogeneity of market impact

So far, I considered the case of a homogeneous market impact parameter $\alpha_j = \alpha$ for all assets (except for cash). In the following, I relax this assumption and explore to what extent this improves the model performance. My main goal here is to show, by proof-of-concept, that there are simple example cases that lead to a superior performance. For this purpose, I assume that the two relevant asset classes (1 and 5), are ten times less liquid compared to the other asset classes. Figure 5.11 and Figure 5.12 compare the model performances for homogeneous (top panels) versus heterogeneous market impact parameters (bottom panels) for the cases $\gamma = 0$ (Figure 5.11) and $\gamma = 20$ (Figure 5.12). Each column in the figures corresponds to an initial shock on a specific asset class. For all asset classes, I see that ROC curves for heterogeneous market impact are superior to those for homogeneous market impact.

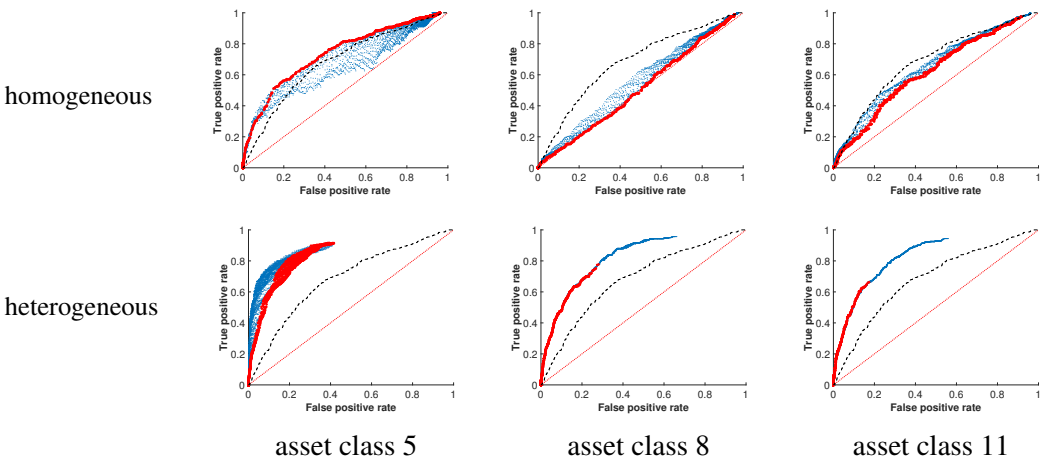


Figure 5.11: Comparison between ROC curves of model with homogeneous market impact parameters (top panels) versus model with heterogeneous market impact (bottom panels). Each column corresponds to an initial shock on a specific asset class (see Table 5.1 for the classification). Here I consider $\gamma = 0$.

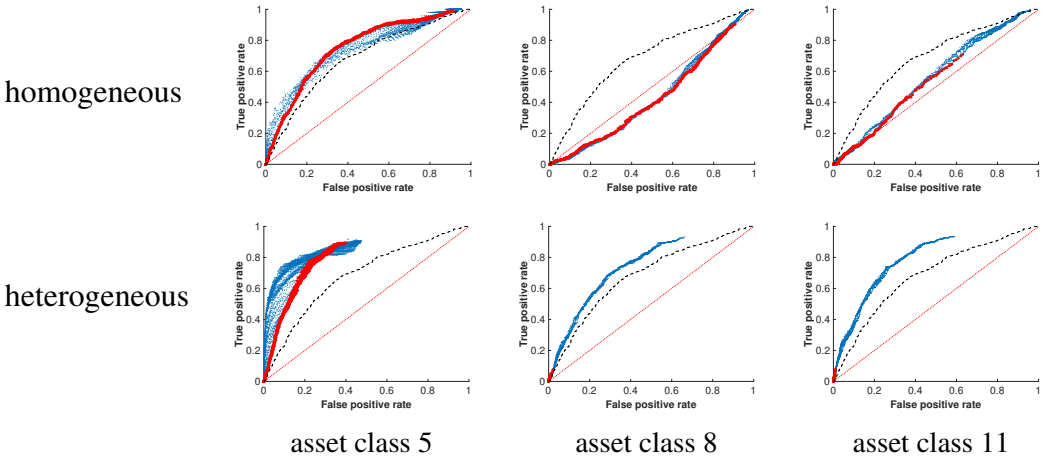


Figure 5.12: Comparison between ROC curves of model with homogeneous market impact parameter (top panels) versus model with heterogeneous market impact parameter (bottom panels). Each column corresponds to an initial shock on a specific asset class (see Table 5.1 for the classification). Here I consider $\gamma = 20$.

Overall, the performance of the model improves for my specific choice of heterogeneity in α. In both figures, the ROC curves in the bottom panels lie above the corresponding ones in the top panels, even if I assume the initial shock affects more liquid asset classes. See, for example, the results for an initial shock on asset class 8 (loans to individuals, center column) and on asset class 11 (held-to-maturity securities, right column).

5.3.2.2 The impact of multiple rounds of asset liquidations

So far I only included one round of asset liquidations in my stress testing model. Here I explore to what extent accounting for multiple rounds affects the predictive performance of the model. In particular, I contribute to the debate as to whether one should consider multiple rounds of asset liquidations. For example, Greenwood et al. (2015) and Duarte and Eisenbach (2015) find that most of the contagion process is captured by the first liquidation round. On the other hand, Cont and Schaanning (2017) argue that this approach may lead to an underestimation of systemic risk.

I look at the effect of multiple rounds of liquidations along two different dimensions: first, I check whether it increases the TPR (the number of bank failures that

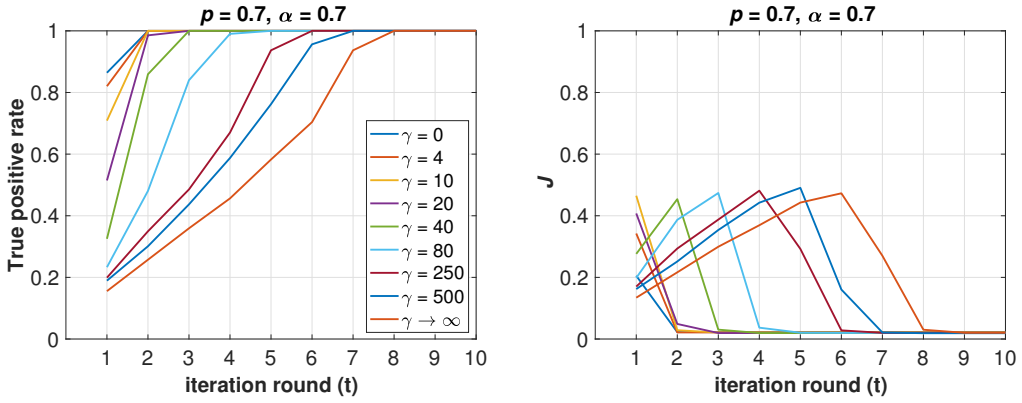


Figure 5.13: (a) The number of bank failures that is correctly identified by the model (normalized to the actual number of bank failures), as a function of γ and iteration round (t). (b) The corresponding value of J as a function of γ and iteration round (t). The results correspond to a combination of p and α in the restricted parameter range ($p = 0.7$ and $\alpha = 0.7$).

the model correctly predicts) of the different models. To this end, I plot the TPR, for several values of γ, as a function of the iteration round (t) in Figure 5.13(a).[10] In line with the results of Cont and Schaanning (2017), I find that the model underestimates the number of banks that fail when only one round of asset liquidations is included. As expected, the number of bank failures increases with t for all values of γ and all models approach $TPR = 1$ at some t. For example, the model with $\gamma = 0$ correctly predicts the defaulted banks (TPR $= 1$) at $t = 2$. Note that, as γ becomes larger, the model needs to include additional rounds of asset liquidations to correctly classify all failed banks. For example, the model with $\gamma \to \infty$ reaches TPR $= 1$ at $t = 8$.

Second, I investigate to what extent the increased TPR from Figure 5.13(a) corresponds to higher values of J. In this respect, higher (lower) values of J would suggest higher (lower) accuracy, i.e. the TPR increases faster (slower) with t compared to the FPR. To this end, I plot J as a function of the iteration round (t) in Figure 5.13(b). Here I again observe that the dynamics of J depend on the assumed

[10] In the plot, I only show the results for $p = \alpha = 0.7$ that correspond to an example from within the restricted range of parameters that I considered ($p, \alpha \in [0.6, 1]$). I note that the results are robust for different parameters within (but also outside of) this range. For the sake of illustration, in Appendix A.6, I show the same plot for $p = \alpha = 0.5$.

liquidation dynamics: for small values of γ (e.g. $\gamma = 0, 4, 10, 20$), I see that the value of J decreases over t. In other words, when banks are aggressive leverage targeters, the model becomes most accurate at $t = 1$; in this case, including additional rounds of asset liquidations reduces the accuracy of the model. This is in line with the findings of Greenwood et al. (2015) and Duarte and Eisenbach (2015).

For larger values of γ, however, Figure 5.13(b) shows that the peak of J shifts further to the right. In other words, when banks are less aggressive to leverage targeters, considering multiple rounds of asset liquidations is more favourable. In this case, bank defaults propagate more slowly through the system, such that additional rounds of asset liquidations are necessary. I should note, however, that J decreases for relatively large t, such that including too many rounds of asset liquidations will always reduce the performance of the stress testing model.

5.4 Conclusion

I studied a network model of price-mediated contagion via fire sales that interpolates between two models that were previously considered in the literature. My model contains a free parameter, γ, which determines how aggressively banks target their leverage ratios (and thus sell assets during distress). I tested the predictive accuracy of the model on empirical data for U.S. bank failures during the recent global financial crisis.

My analysis has important implications for the application of macroprudential stress tests. To analyse the stability of financial networks due to price-mediated contagion, one needs to make assumptions regarding the behavioural response of banks (in term of liquidation behaviour) during the stress scenario. In this chapter, I showed that it is important to consider a range of assumptions in relation to the behaviour of banks. I provided a framework to do this in a structured manner by means of a free model parameter (γ) and illustrated how the best performing value of γ depends on the choice of the other model parameters, in particular the size of the initial shock and the market impact parameter. Moreover, I showed that the overall model predictive performance strongly depends on: 1) the type of shock

being imposed (systematic versus idiosyncratic), 2) the asset class-specific market impact parameters, and 3) the number of liquidation rounds considered in the stress testing model.

My findings suggest several interesting paths for future research. First and foremost, I think it is of utmost importance to perform similar analyses on other datasets. This chapter specifically focuses on the behaviour of U.S. banks during the global financial crisis. Future research should therefore explore to what extent the findings vary for different countries and/or different time periods. Another important follow-up question is whether individual banks follow different behavioural strategies, i.e., considering different values of γ for each bank. This might further increase the predictive accuracy of the proposed stress testing model.

Chapter 6

On System-Wide Stress Tests

6.1 Introduction

"The Whole is Greater than the Sum of its Parts."'

– Aristotle

Most of the literature to date has only looked at systemic risk within one financial sector in isolation. In this chapter, I look at common asset holdings between UK banks, UK open-ended investment funds, and UK (both unit-linked and non unit-linked) insurance companies. My dataset consists of portfolio holdings of marketable assets such as equities and debt securities (corporate and government bonds) at the instrument level for the period Q1 2017. The equities and debt securities coverage accounts for 50.5% activity of the UK open-ended funds, 80% activity of the UK banks' regulated by the Prudential Regulation Authority, and 84.6% activity of the UK insurance companies. My aim is to study the extent to which common asset holdings across these sectors contribute to systemic risk.

To this end, I first build a bipartite network of common asset holdings where institutions are connected to the assets they hold. I then show that there are portfolio similarities across the different sectors. Furthermore, I consider a stress test model, where I assume that different sectors are subject to different constraints. In particular, banks and non unit-linked insurers are forced to liquidate (some of) their assets to comply with regulatory constraints. Meanwhile, funds and unit-linked insurers are obliged to sell their assets to meet investor redemptions. Following Greenwood

et al. (2015) and Fricke and Fricke (2020), I look at three different measures of systemic losses: (i) fire sale losses of the system, (ii) indirect vulnerability and (iii) systemicness of each institution and/or sector.

My main findings are as follows: first, I find the importance of considering multiple financial sectors in the analysis of systemic risk. In particular, I show that ignoring common asset holdings between different sectors may result in an underestimation of systemic losses by 47% on average. Second, I conduct a systemic stress test on UK banks and non-banks under different types of initial shocks. In most instances, I find that fire sale losses resulting from assets liquidations are higher than direct losses from initial shocks. Moreover, I look at the case when institutions maintain their portfolio weights (pro-rata liquidation) vs. the case when institutions prefer to sell their most liquids assets first (waterfall liquidation). I show that the pro-rata liquidation approach always yields a higher level of systemic risk. However, I note that the waterfall liquidation may produce a higher spillover effect (indirect vulnerability) for an institution or a sector that chooses not to liquidate any of its assets during distress. Finally, I look at the importance of each institution in the network, based on their systemicness and vulnerability. Third, I show that portfolio similarity measures are useful to explain more variability in the stress simulation results. In particular, I find that larger and more diversified institutions will become less (indirectly) vulnerable.

The remainder of the section is organised as follows: in section 6.2, I describe the contagion model, statistical characterization of dataset and experimental setup. In section 6.3, I present and discuss the results. Finally, I discuss my conclusions in section 6.4.

6.2 Methodology

6.2.1 Model

In the previous chapters, I have mainly looked at the systemic risk in banking networks. In the following, I study the contagion due to common asset holdings between banks and non-banks financial intermediaries. The regulatory constraint for

banks was previously discussed in chapter 3. In the following, I describe the responses of funds and (both unit-linked and non unit-linked) insurance companies to a financial shock.

6.2.1.1 Funds

The balance sheet of a fund is illustrated in Figure 6.1. Unlike banks, funds do not need to comply with any regulatory constraints. However, as it was empirically shown in Czech and Roberts-Sklar (2019), funds are pro-cyclical and liquidate their assets to meet investor redemption at periods of stress.

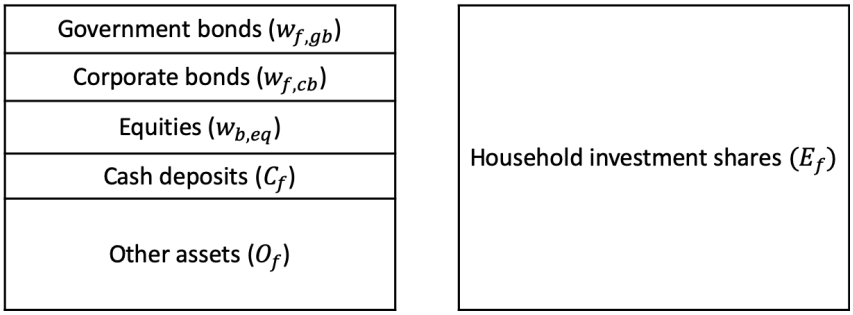


Figure 6.1: Fund's balance sheet. The left panel is the asset side, while the right is the liability side of the balance sheet.

Following Baranova et al. (2017) and Fricke and Fricke (2020), the amount of assets that fund f liquidates is:

$$\Pi_f^1 = \sigma_f \phi_f^1 A_f^1, \tag{6.1}$$

where ϕ_f^1 is the loss that fund f receives (see Equation 3.6), and σ_f is the fund-flow performance sensitivity parameter, which is the share of assets that investors will redeem following losses of 1%.

6.2.1.2 Unit-linked insurers

The balance sheet of a unit-linked insurer is illustrated in Figure 6.2. The business models of unit-linked insurers and funds are similar, as they both pool policyholders/investor funds and invest them in financial assets. However, unlike funds, unit-linked insurers tend to have longer-term horizons. This means that their policyholders may be better able to accept shorter-term portfolio losses, hence unlikely to

redeem their funds during distress. Nevertheless, they are given an option to switch their investments between different asset classes. In fact, in its recent survey, Bank of England (2016) observe that some unit-linked policyholders decide to de-risk their investment in response to falling prices of risky assets.

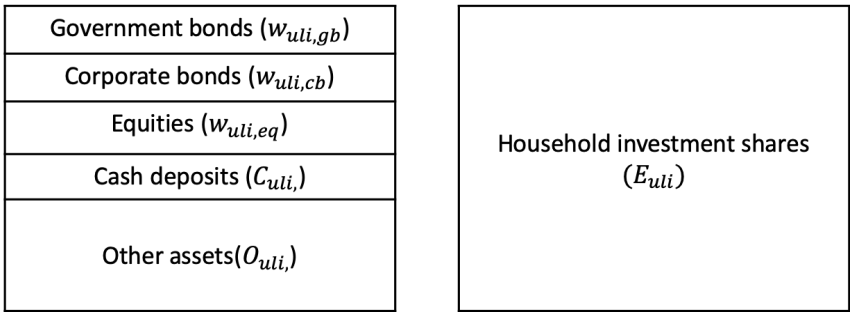


Figure 6.2: Unit-linked's balance sheet. The left panel is the asset side, while the right is the liability side of the balance sheet.

Following Baranova et al. (2019), the amount of total assets that unit-linked insurer *uli* needs to liquidate is given by:

$$\Pi^1_{uli} = \sigma_{uli}\phi^1_{uli}(A^1_{uli,eq} + A^1_{uli,cb}), \tag{6.2}$$

where σ_{uli} is the policyholder switching sensitivity parameter, which is the share of assets that policyholders will switch following losses of 1%. Note that unit-linked insurers will only liquidate risky assets, and $A_{uli,eq}$ and $A_{uli,cb}$ are the total portfolio holdings of equities and corporate bonds.

6.2.1.3 Non unit-linked insurers

Finally, I look at the case of non unit-linked insurers, and illustrate their balance sheet in Figure 6.3. As shown in the figure, non unit-linked insurers are similar to banks, in a way that they both hold capital and have to comply with some regulatory constraints.

Following Aikman et al. (2019), I assume that non unit-linked insurers target their solvency ratio, which is defined as:

$$SR^0_{nli} = \frac{E^0_{nli}}{SCR^0_{nli}}, \tag{6.3}$$

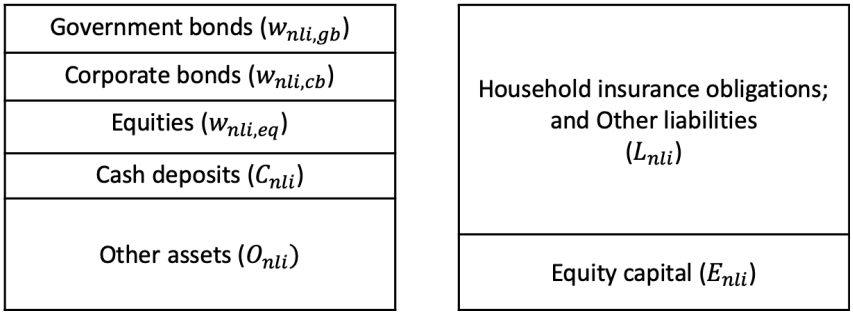


Figure 6.3: Non unit-linked insurer's balance sheet. The left panel is the asset side, while the right is the liability side of the balance sheet.

where SCR^0_{nli} is the regulatory solvency capital requirement:

$$SCR^0_{nli} = kn_{nli} + (A^0_{nli,eq} + A^0_{nli,cb} + O_{nli})km_{nli}. \tag{6.4}$$

Note that $A_{nli,eq}$ and $A_{nli,cb}$ in Equation 6.4 are the total portfolio holdings of equities and corporate bonds. This would imply that non unit-linked insurers will only liquidate risky assets. Meanwhile, kn_{nli} and km_{nli} are the capital charges for non-market and market risks.

Following the Solvency II regulation,[1] non unit-linked insurers are reasonably well hedged in general. For example, when non unit-linked insurers sell corporate bonds, they will also lose some of the hedging benefits (or matching adjustment) in their balance sheet. This means that they would see a decrease in their liabilities, and consequently an increase in their equity and solvency ratio.

I therefore assume that non unit-linked insurers attempt to maximise the value of their equity. To this end, the total assets that they would sell are computed using a measure of post-shock elasticity that is collected by the Prudential Regulatory Authority. Suppose that $E\prime^1_{nli}$ is the new equity, and $SR\prime^1_{nli}$ is the new solvency ratio that are computed using the elasticity measure. As I assume that non unit-linked insurers target their solvency ratio: SR^0_{nli}, the amount of SCR_{nli} that the insurer

[1] https://www.bankofengland.co.uk/prudential-regulation/key-initiatives/solvency-ii

needs to reduce can be computed as:

$$\Delta SCR^1_{nli} = E\prime^1_{nli} \left(\frac{SR^0_{nli} - SR\prime^1_{nli}}{SR^0_{nli} \times SR\prime^1_{nli}} \right). \tag{6.5}$$

The total risky assets that the insurer sells it therefore:

$$\Pi^1_{nli} = \frac{\Delta SCR_n li^1}{km_{nli}}. \tag{6.6}$$

6.2.2 Data

In this chapter, I use granular equity and debt security holdings of the seven UK banks[2] that took part in the 2017 annual cyclical scenario, UK open-ended investment funds and UK (both unit-linked and non unit-linked) insurance companies for Q1 2017 reporting period. Each asset in my dataset is identified by an ISIN. Let us start by describing the sources of these datasets. Below I provide details of the three datasets used.[3]

- Banks: I use proprietary data submitted to the Prudential Regulatory Authority by the seven UK banks that took part in the 2017 annual cyclical stress test. Banks should report the exposure amount in the currency of the security at ISIN level.

- Open-ended investment funds: I extract from Morningstar voluntarily reported data on open-ended investment funds that are domiciled in UK. In particular, I use granular data on portfolio holdings that include holding type and unique identifiers such as ISINs. I also use data on total net assets and on the funds investment profiles.

[2]The 2017 stress test covered seven major UK banks and building societies (hereafter 'banks'): Barclays, HSBC, Lloyds Banking Group, Nationwide, The Royal Bank of Scotland Group, Santander UK and Standard Chartered.

[3]Barucca et al. (2020) use a similar type of dataset to perform an extensive study on the network of common asset holdings across financial sectors. Note that they consider an aggregated version of my data, where financial assets are grouped according to their issuers. Furthermore, they look at the network between European investment funds, UK banks and UK insurance companies for Q1 2016 reporting period, while I focus on the network between UK domiciled financial institutions for Q1 2017 reporting period.

- Insurance Companies: I sample granular line-by-line asset data from Prudential Regulatory Authority regulated UK insurance companies subject to the Solvency II directive. My data includes unique identifiers, such as ISINs and LEIs of counterparties, as well as categorisation of assets into Complementary Identification Code types. For the purpose of this analysis I consider both unit-linked and non unit-linked portfolios.[4]

I note that the data described above is non-public. Therefore, I only present results in anonymised or aggregated format.

6.2.2.1 Network of common asset holdings

I combine the datasets for different financial sectors and construct a network of common asset holdings. In the following, I describe the properties of the corresponding network. Table 6.1 summarises the properties of each financial sector. As shown in the table, the total holdings in the whole network is £2.04 trillion. Funds account for approximately 40% of the total holdings, which is twice as much as the contribution of banks or insurers. This is due to a large number of funds ($n = 1865$) that exist in the network. In fact, as shown in the table, the average size of each fund is relatively small. For instance, funds' average strength is only £0.43 billion, much smaller than the average strength of banks (£60.04 billion). The same is true also for the their average connection, as shown by the average degrees reported in Table 6.1. Overall, I find that the network is very sparse, with a density of only 0.30%.

Data for Q1 2017	Banks	Funds	ULI	NLI	All firms
Number of entity	7	1865	31	20	1923
Total holdings	420.27	805.15	461.14	356.58	2043.10
Average strength	60.04	0.43	14.88	17.83	1.06
Average degree	1427	88	1499	1321	127
Density (%)	3.35	0.20	3.52	3.10	0.30

Table 6.1: Summary properties of each **financial sector** in the network of common asset holdings. ULI corresponds to unit-linked, while NLI to non unit-linked insurance companies. Average strength and total holdings are presented in £bn.

[4]It is possible for an insurance company to be linked and non unit-linked at the same time and thus be represented by two separate nodes in my analysis.

I present the summary properties of each asset class in Table 6.2. As shown in the table, there are 42611 instruments in total, where each of them belongs to a particular asset class: equities, corporate bonds or government bonds. In term of the size, equities are the largest asset class in the network, accounting for up to 50% of all assets in the network. Additionally, I observe that the average strength (average degree) of government bonds is the largest (smallest) compared to other sectors. This implies that the individual investment in government bonds is relatively higher compared to that in equities and corporate bonds. In addition to these aggregate summary properties, I also plot the degree distribution of each institution and each asset in Figure 6.4. From the figure, I observe the variability of degree distributions among institutions and assets.

Data for Q1 2017	Equities	Corp bonds	Gov bonds	All assets
Number of entity	19847	17103	5661	42611
Total shares	1060.80	413.70	568.61	2043.10
Average strength	53.45	24.19	100.44	47.95
Average degree	8.05	3.80	3.52	5.74
Density (%)	0.42	0.20	0.52	0.00

Table 6.2: Summary properties of each **assets class** in the network of common asset holdings. Average strength and total holdings are presented in £bn.

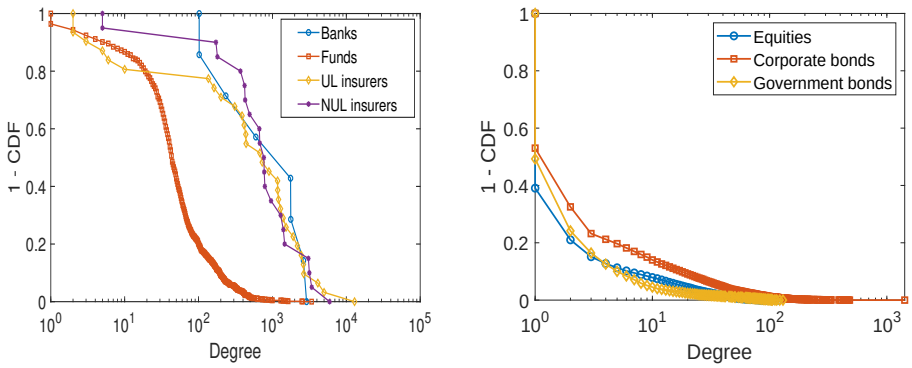


Figure 6.4: Degree distribution of each financial institution (left) and each financial asset (right).

6.2.2.2 Holdings across sectors and asset classes

I present the portfolio holdings of each financial sector across different asset classes in Table 6.3. Overall, I find that the relative balance sheet composition varies across

sectors. As shown in the table, most of the portfolio holdings of banks and non unit-linked insurers consist of bonds, while funds and unit-linked insurers hold mostly equities. This composition results in the variation of relative losses that each sector may receive following an initial shock to a particular asset class.

	All assets	Equities	Corp bonds	Gov bonds
Banks	420.27	33.56	85.37	301.34
Funds	805.15	649.48	93.38	62.29
ULI	461.14	318.98	48.26	93.90
NLI	356.58	58.81	186.68	111.09
All sectors	2043.10	1060.80	413.70	568.61

Table 6.3: Aggregate total holdings (in £bn) for each financial sector across different asset classes.

6.2.2.3 Portfolio similarity

I first discuss the average portfolio similarity across different pairs of institutions in specific sectors. I present the values in Table 6.4 and Table 6.5, for the binary and weighted measure respectively. First and foremost, I find that there are portfolio similarities across the different sectors. Moreover, with the exception of the binary similarity across funds, I find that the portfolio similarities across the same sector are higher compared to those across the different sectors. Additionally, I observe that the binary and weighted measure may produce different results. For example, Table 6.4 shows that the result across non unit-linked insurers is higher compared to that across banks, suggesting that non unit-linked insurers have a larger number of assets in common. However, Table 6.5 shows that the opposite is true, indicating that banks have more portfolio weight in common.

	Banks	Funds	ULI	NLI
Banks	114.33	4.79	44.70	56.37
Funds		3.50	18.27	14.99
ULI			180.81	160.69
NLI				189.46

Table 6.4: Average binary portfolio similarity across different sub-networks corresponding to different pairs of sectors in the common asset holdings network.

	Banks	Funds	ULI	NLI
Banks	0.19	0.02	0.08	0.11
Funds		0.29	0.07	0.03
ULI			0.21	0.11
NLI				0.18

Table 6.5: Average weighted portfolio similarity across different sub-networks corresponding to different pairs of sectors in the common asset holdings network.

Second, I look at the overall average portfolio similarity across all institutions in the network. In Table 6.6, I present the results for the binary and weighted measure. In average, I find that each institution has 4.42 number of assets in common and 28% similarity in their portfolio weights.

	Mean	Std	Max	Min
MeanBinSimilarity	4.42	6.41	74.78	0.00
MeanCosSimilarity	0.28	0.20	0.52	0.00

Table 6.6: Average binary and weighted portfolio similarity over all institutions in the common asset holdings network.

6.2.3 Experimental setup

6.2.3.1 Sector-specific constraint

In the following, I present the aggregate statistics of banks' leverage ratio and non unit-linked insurers' solvency ratio. Furthermore, I describe the calibration of fund-flow performance sensitivity and unit-linked policyholders' switching parameters.

Banks In Table 6.7, I present the aggregate statistics of banks' leverage ratio. As shown in the table, the average leverage ratio of banks in my datasets is 19.50, which is below the 33.33 UK minimum leverage requirement.[5]

Funds I consider the fund-flow sensitivity parameters across the different categories of funds that has been calibrated previously in Baranova et al. (2017). Specifically, they have run a panel regression on Morningstar European fund-level monthly data on TNA and Estimated Net Flows from January to September 2016. I present these parameters in Table 6.8,

[5] In this thesis, the leverage ratio is defined as the ratio between the total assets and the capital, that is $\frac{A_i}{E_i}$ (see Equation 3.5).

	Total all assets (£bn)	Leverage ratio (%)
Average	803.57	19.50
Std	530.97	150.90
25th percentile	303.00	23.26
Median	677.00	19.23
75th percentile	1207.00	17.54

Table 6.7: Aggregate statistics of banks' leverage ratio and total assets. Note that total assets here is different from the total portfolio holdings, as it also consists cash reserves, derivatives and interbank assets.

Category of funds	Fund-flow sensitivity parameter
Allocation	0.20
Commodities	0.10
Convertibles	0.43
Equity	0.09
Fixed income	0.52

Table 6.8: Fund-flow sensitivity parameter across different categories of funds, as was previously calibrated in Baranova et al. (2017).

Unit-linked insurers In terms of the sector-specific constraint of unit-linked insurers, I consider an investor switching parameter that has been previously used in Baranova et al. (2019), which is based on the survey of Bank of England (2016). In particular, I use $\sigma_{uli} = 0.3$.

Non unit-linked insurers In Table 6.9, I present the aggregate statistics of equity capital and solvency capital requirement (SCR) of non unit-linked insurers in my dataset. Moreover, I calibrate the average capital charge on risky assets as in Aikman et al. (2019), who assume that the capital charge for risky assets is 50% of total capital requirement, i.e. $km_{nli} = 0.5$.

	Equity capital (£bn)	SCR (£bn)
Average	5.72	3.43
Std	5.92	3.07
25th percentile	1.98	1.35
Median	3.60	2.66
75th percentile	7.90	3.97

Table 6.9: Aggregate statistics of non unit-linked insurers' equity capital and solvency capital requirement (SCR).

6.2.3.2 Initial shocks

I consider two types of initial shock: 1) idiosyncratic shock on each or all asset classes, and 2) regulatory stress test scenario. The latter includes the Comprehensive Capital Analysis and Review (CCAR) stress test scenario of the Federal Reserve Board 2017, and the Bank of England ACS scenario in 2017. Both regulatory scenarios provide the percentage change of each asset class across different jurisdictions. The CCAR scenario covers a broader range of jurisdictions, as it includes 80.6% of assets in my dataset. Meanwhile, the Bank of England scenario includes 74.8% of assets in my dataset, and it focuses on more liquid markets.

Note that the shock of an equity asset in both regulatory scenarios is given in terms of its original price, and therefore can be directly used in my framework. However, the shock of a corporate and government bond is provided in terms of its original yield, and therefore needs to be converted. Suppose dy is a change in the bond yield, the percentage change in its price (dp/p) can be computed as:

$$\frac{dp}{p} = -D * dy\,, \tag{6.7}$$

where D is the modified duration of the bond, that is the measure of its price sensitivity to changes in its yield to maturity.

6.2.3.3 Market depths

Table 6.10 summarised the market depth values that were previously calibrated at asset class level for Q1 2016 reporting period in Barucca et al. (2020). I scale these values to obtain the market depths at individual instrument level.

Asset class	Market depth (£bn)
Equities	338.75
Corporate bonds	55.46
Government bonds	338.75

Table 6.10: Market depth at asset class level for Q1 2016 reporting period that was previously in Barucca et al. (2020).

Suppose δ_J is the market depth of asset class J and S_J^A is the total shares of asset class J held in the network. Let j be an instrument that belongs to class J with

the total shares equal to S_j^A. The market depth of instrument j can be calculated as:

$$\delta_j = \frac{S_j^A}{S_J^A}\delta_J. \tag{6.8}$$

By doing such rescaling, I am assuming that an asset with a larger (smaller) value of total shares will have a larger (smaller) value of market depth, therefore the asset is more liquid (illiquid). For example, I see from Table 6.10 that the market depth of equities is £338.75bn, i.e. δ_{EQ} = £338.75bn. Let suppose there are only two equity assets in my network, eq_1 and eq_2. If the total shares of eq_1 and eq_2 are respectively £10bn and £25bn, i.e. $S_{eq_1}^A$ = £10bn and $S_{eq_2}^A$ = £25bn, I will then have S_{EQ}^A = £10bn + £20bn = £35bn. Therefore, the market depth of eq_1 and eq_2 that I will obtain are respectively δ_{eq_1} = £$\frac{10}{35}$ × 338bn = £96.57bn and δ_{eq_2} = £$\frac{25}{35}$ × 338bn = £241.43bn.

I present the results of such rescaling in Figure 6.5, where I plot the distribution of the scaled market depth of each asset. The plot shows that some government bonds seem to be more liquid than equities. This is related to the fact that 35.8% of government bonds in my dataset are based in the U.S., and therefore are extremely liquid. The plot also shows that a few corporate bonds are much more liquid than some equities, which is reasonable if the former are based in the advanced economies while the latter are in the emerging markets.

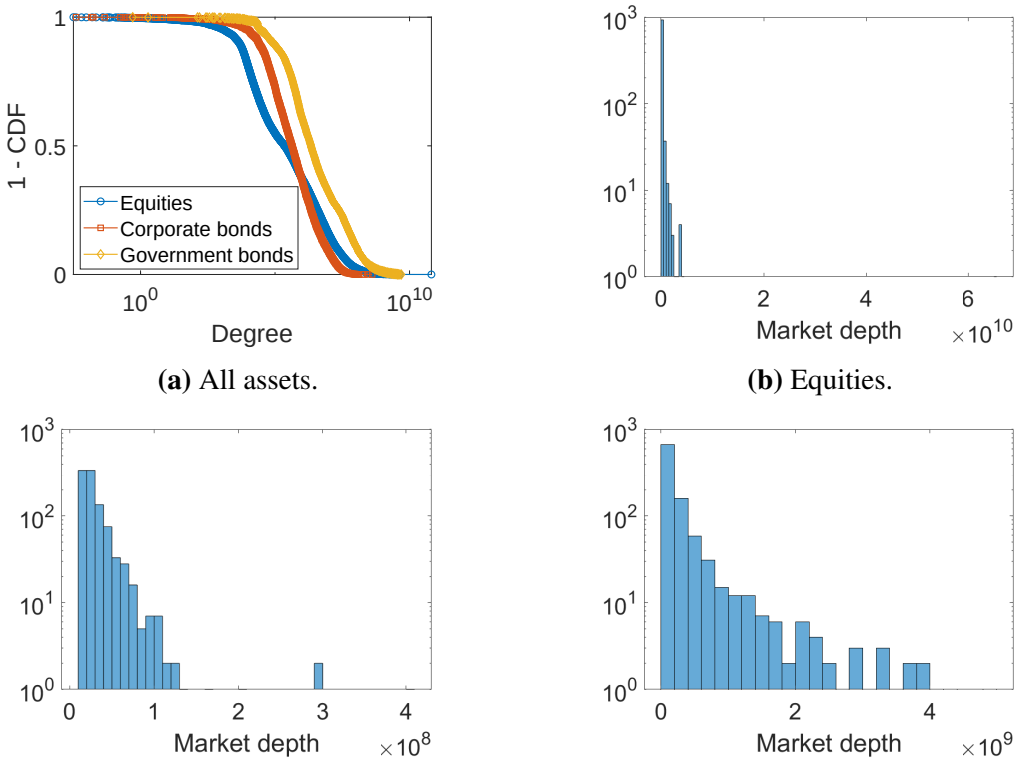
6.2.4 Measuring fire sale spillovers

Following Greenwood et al. (2015) and Fricke and Fricke (2020), I monitor three different measures to quantify the effect of fire sales. Firstly, I look at the *aggregate fire sale losses* that I define as:

$$R^{\text{firesales}} = \sum_i \sum_j \left(w_{ij}^1 - \pi_{ij}^1\right) \times \Psi_j(\beta_j^1). \tag{6.9}$$

It is important to note that this formula only accounts for spillovers losses and ignores direct losses:

$$R^{\text{direct}} = \sum_i \sum_j w_{ij}^0 p_j^0 \tag{6.10}$$



(a) All assets. **(b)** Equities.

(c) Corporate bonds. **(d)** Government bonds.

Figure 6.5: Distribution of the scaled market depth for each asset. Top left: CCDF of the market depth distribution. Top right: histogram of equities, bottom right: government bonds, bottom left: corporate bonds.

that are incurred from initial shocks.

Secondly, I look at the *indirect vulnerability* of institution i, which is the spillover effect that i would receive because other institutions liquidate their assets assuming that i does not liquidate any of its assets. Formally, I define the indirect vulnerability as follows:

$$R_i^{\text{vulnerability}} = \sum_j \left(w_{ij}^1 - \pi_{ij}^1\right) \times \Psi_j(\beta_j^1) \quad \text{where} \quad \pi_{ij}^1 = 0 \quad \text{and} \quad \pi_{kj,k\neq i}^1 \geq 0. \tag{6.11}$$

Finally, I can calculate the marginal contribution of i to the aggregate fire sale losses. Specifically, I assume that i is the only institution that would liquidate its

asset. To this end, I define the *systemicness* of i as:

$$R_i^{\text{systemicness}} = \sum_k \sum_j \left(w_{kj}^1 - \pi_{kj}^1 \right) \times \Psi_j(\beta_j^1) \quad \text{where} \quad \pi_{ij}^1 \geq 0 \quad \text{and} \quad \pi_{kj,k\neq i}^1 = 0 \tag{6.12}$$

In line with these definitions, I can also define aggregate indirect vulnerability and systemicness measures for each sector.

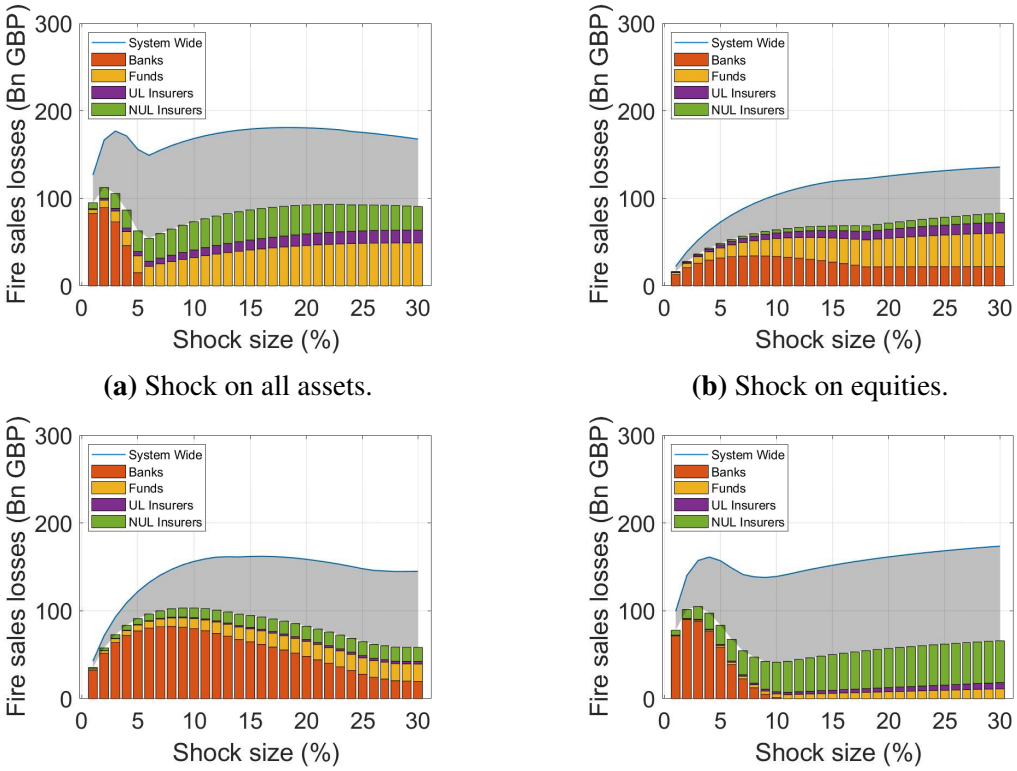
6.3 Results

In the following, I present and discuss the results obtained from modelling fire sale contagion across different sectors. In particular, I look at three different measures of fire sales: (i) aggregate fire sale losses, (ii) institution's (sector's) indirect vulnerability, and (iii) institution's (sector's) systemicness. As explained in subsubsection 6.2.3.2, I consider two types of initial shock: (i) idiosyncratic shock on asset class(es), and (ii) regulatory shock scenarios.

6.3.1 The whole is different from the sum of its parts

Let me start by discussing the differences between modelling contagion across sectors vs. within each sector separately. To obtain the aggregate fire sales losses of the former exercise, I simply run the model on the complete network of common asset holdings that consists of banks, funds and (both unit-linked and non unit-linked) insurance companies. Meanwhile, the results of the latter can be measured by running the model on each sub-network separately, where each sub-network consists only of financial institutions within the same sector. Note that the total liquidated assets in both exercises are exactly the same. The important question is, however, whether the total losses are also identical. In other words, I want to look at whether the whole is the sum of its parts.

Figure 6.6 shows the results of the two exercises. The stacked bar charts in the figure corresponds to the cumulative results for each sub-network, while the line plot is the results for the complete network. Furthermore, the grey shadow area is the differences between the two exercises, which implies that it represents the amount of losses that is due to the common asset holdings across different sectors.



(a) Shock on all assets.

(b) Shock on equities.

(c) Shock on corporate bonds.

(d) Shock on government bonds.

Figure 6.6: The whole is different from the sum of its parts. Stacked bar charts corresponds to the cumulative aggregate fire sale losses from modelling the contagion for each sub-networks separately, where each sub-network consists only of institutions within the same sector. A blue line corresponds to results for the complete network, where the network consists of institutions across multiple sectors. The differences between the two results is shown as a grey shadow area.

The figures illustrate that there are large differences between the two results. More importantly, it suggests that ignoring common asset holdings between different financial sectors can result in an underestimation of systemic risk. In Table 6.11, I compute the averages of the underestimation over different sizes of idiosyncratic shocks ($p_j \in \{0, 0.01, 0.02, ..., 0.3\}$). The table shows that the average systemic risk underestimation is around 47%, and it can reach up to 70%.

6.3.2 A systemic stress simulation of the UK financial system

In the previous section, I conducted stress simulations on the UK financial system by applying idiosyncratic shocks on asset class(es) and considering the pro-rata liquidation approach. I then showed the importance of considering multiple financial

Shock on	Mean	Std	Max	Min
All assets	50	8	64	22
Equities	39	5	44	26
Corp. bonds	41	12	60	24
Gov. bonds	60	10	70	40
Total	47	3	70	24

Table 6.11: The amount of systemic risk underestimation (in %) for ignoring the common asset holdings across different financial sectors. Results are computed over different sizes of idiosyncratic shock ($p_j \in \{0, 0.01, 0.02, ..., 0.3\}$) for shock on different asset class(es).

sectors in the analysis. In the following, I extend the analysis by taking regulatory stress scenarios and waterfall liquidation into account. In addition to aggregate fire sale losses, I also look at indirect vulnerability and systemicness of each institution (sector). Furthermore, I provide a map to the most systemic and the most vulnerable institutions in the system.

6.3.2.1 Aggregate fire sale losses

The regulatory stress scenario. I first look at aggregate fire sale losses for the case of regulatory stress scenarios. In particular, I present the results for the Bank of England (BoE) scenario in Table 6.12, while those for the Federal Reserve Board (FRB) CCAR scenario in Table 6.13. Both tables show that the aggregate fire sale losses are larger than the direct losses. For example, in the case of the BoE scenario, I observe the fire sale losses of 5.35% in correspondence to the direct losses of 3.63%. Note that the former is losses due to the contagion only, and exclude those resulting from the initial shock.

	Pro-rata liquidation	Waterfall liquidation
Direct losses	3.62% (74.01 bn)	
Total sales	2.60% (53.03 bn)	
Fire sale losses	5.35% (109.31 bn)	3.69% (75.41 bn)

Table 6.12: Aggregate direct (due to initial shock) and fire sale (due to contagion only) losses for the Bank of England stress scenario.

Second, I find from Table 6.12 and Table 6.13 that the aggregate fire sale losses for the pro-rata liquidation are always larger than those obtained for the waterfall case. For example, the losses for the FRB CCAR scenario is 8.66% for the former,

	Pro-rata liquidation	Waterfall liquidation
Direct losses	3.72% (76.10 bn)	
Total sales	9.39% (191.88 bn)	
Fire sale losses	8.66% (176.86bn)	6.60% (122.93 bn)

Table 6.13: Aggregate direct (due to initial shock) and fire sale (due to contagion only) losses for the Federal Reserve Board CCAR stress scenario.

while only 6.60% for the latter. This result is due to the fact that institutions also sell their illiquid assets during the pro-rata liquidation, which then results in a more severe price impact.

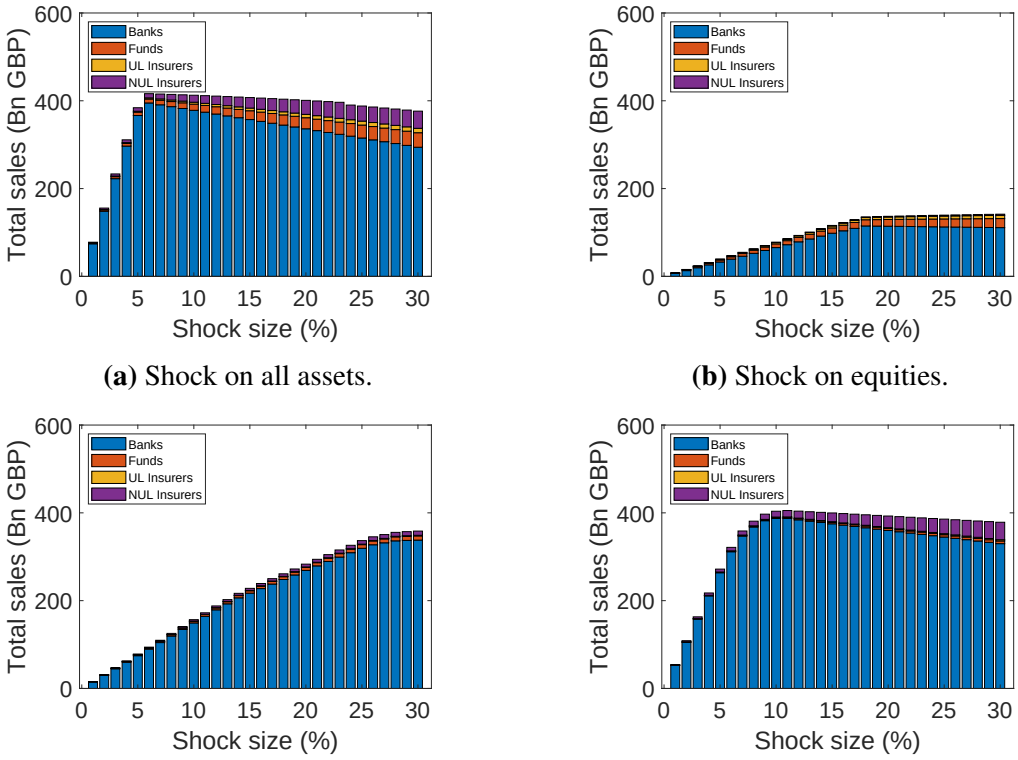
Finally, I observe that direct losses for both scenarios are relatively similar, while their fire sale losses are not. For example, the direct losses for the BoE and FRB CCAR scenario are 3.62% and 3.72% respectively, with a difference of only 0.1% (£2 bn) direct losses between the two. Meanwhile, the corresponding fire sale losses for the pro-rata case are 5.35% and 8.66% respectively, with a difference of 3.31% (£67.55 bn) fire sale losses between the two. This result corresponds to the type of assets being shocked in the two scenarios. For example, the fire sales losses are higher for the FRB CCAR case because the scenario covers a larger number of illiquid assets.[6]

The idiosyncratic stress scenario. The previous finding suggests that fire sale losses do not only depend on the size of initial shock, but also on the type of depressed assets. In the following, I discuss the results for the case of idiosyncratic shocks on a particular asset class.

Let's start by looking at the total amount of liquidated assets in Figure 6.7. The figure shows that the amount varies across different financial sectors and types of shock. For example, banks become the sector that always liquidate the largest amount of assets. This result is due to the sector-specific constraint that was previously described in subsection 6.2.1. Banks, for instance, have the strongest constraint since they will need to target their leverage ratio. Moreover, Figure 6.7 shows that banks liquidate a larger amount of assets when the shock is imposed on bonds compared to when the shock is imposed on equities. This is due to the relative

[6] See the coverage of each scenario in subsubsection 6.2.3.2.

balance sheet composition of each sector across different asset classes, as was previously presented in in Table 6.3. The balance sheet of banks, for instance, consists mostly of government bonds.



(a) Shock on all assets.

(b) Shock on equities.

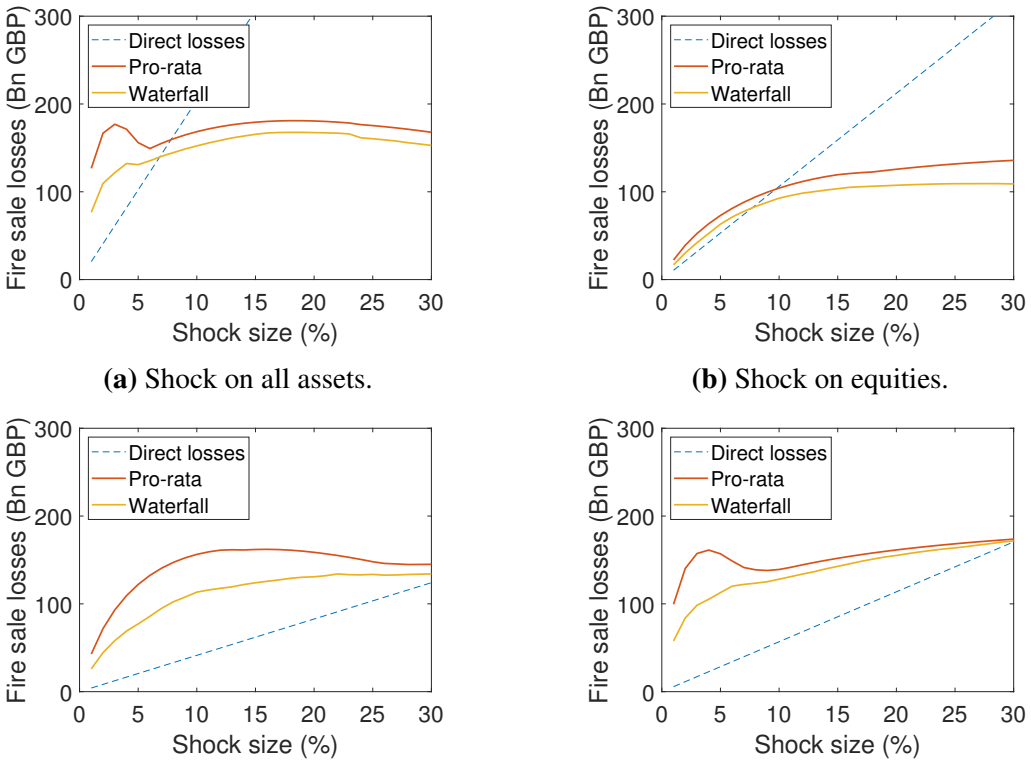
(c) Shock on corporate bonds.

(d) Shock on government bonds.

Figure 6.7: Total volume of liquidated assets across different sectors for different types of idiosyncratic shocks.

Furthermore, Figure 6.8 shows the corresponding aggregate fire sale losses, for the pro-rata and waterfall liquidation. Interestingly, the figure shows the existence of inverted u-shaped curves, when the shock is imposed on government bonds and all assets. The reason for this behaviour is that the amount that institutions liquidate increases with the size of the shock as long as they have enough assets to liquidate. When this happens, some institutions do not longer experience fire sale losses simply because they are left with no available assets to sell. So, as the shock increases, institutions move from a small shock regime where their losses are mostly due to fire sale devaluation to a large shock regime where their losses are dominated by the shock. This is the reason for of the non-monotonicity observed in Figure 6.8a and

6.8d.



(a) Shock on all assets.

(b) Shock on equities.

(c) Shock on corporate bonds.

(d) Shock on government bonds.

Figure 6.8: Aggregate fire sale losses for different types of idiosyncratic shocks. Red line corresponds to the losses for the pro-rata liquidation, while yellow line refers to those for the waterfall liquidation. Blue dashed-line is the corresponding aggregate direct losses.

Finally, Figure 6.8 also shows the comparison of fire sale losses between the pro-rata and waterfall liquidation. For all types of idiosyncratic shocks, I observe that the losses for the pro-rata liquidation are always more severe than those for the waterfall liquidation. This finding is consistent with the results for the BoE and FRB CCAR stress scenario reported in Table 6.12 and Table 6.13.

6.3.2.2 Indirect vulnerabilities

My previous results indicate that the pro-rata liquidation leads to the highest aggregate losses. In the following, I would like to see whether this is the case for all institutions and sectors. In particular, I compare the indirect vulnerability resulting from the two liquidation approaches. The idea is to measure the spillover effect that an institution (sector) receives because other institutions liquidate their assets,

assuming that the institution (sector) to be passive.

In Figure 6.9, I present the indirect vulnerability of each institution when the other institutions follow the pro-rata vs. the waterfall liquidation approach. Each dot in the figure corresponds to the calculation of the vulnerability of an institution. If the pro-rata liquidation is worse than the waterfall liquidation for all firms, all dots should lie below the black dashed diagonal line. Figure 6.9 shows that this is not the case. In fact, I observe that several institutions lie above the diagonal line, suggesting that they are more vulnerable when other institutions use the waterfall approach.

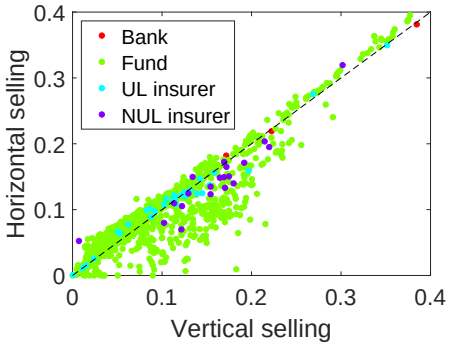


Figure 6.9: Indirect vulnerability that institutions receive when other institutions consider the pro-rata vs. waterfall liquidation approach, for the case of 5% initial shock on all assets (relative to the total assets of the corresponding institution). Each colour in the plot corresponds to the result for different financial sectors. A dot lies above (below) the black diagonal dashed line will imply that the corresponding institution is more vulnerable when other institutions choose the waterfall (pro-rata) approach.

Furthermore, I look at the case of indirect vulnerability for each sector. In particular, I present the results for different types of idiosyncratic shocks in Figure 6.10 for banks, and in Figure 6.11 for funds. Both figures again show that banks and funds, in general, are more vulnerable if other sectors use the waterfall liquidation approach.[7]

Overall, I find that the waterfall approach may result in more vulnerable institutions (sectors). The intuition behind this result is the following: the prices of liquid assets will fall harder if all other firms prefer to liquidate their most liquid assets.

[7]I observe a similar finding for the case of both unit-linked and non unit-linked insurance companies. See Figure C.1 and Figure C.2 in Appendix.

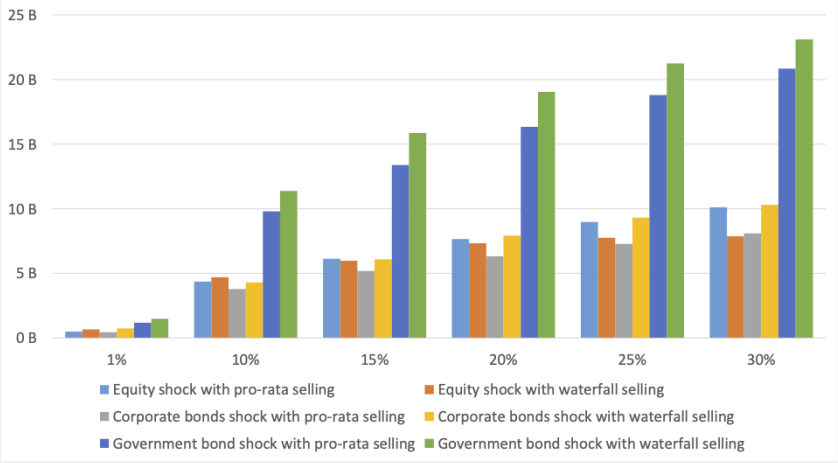


Figure 6.10: Indirect vulnerability of banks for different types of idiosyncratic shocks, when other sectors follow pro-rata vs. waterfall liquidation.

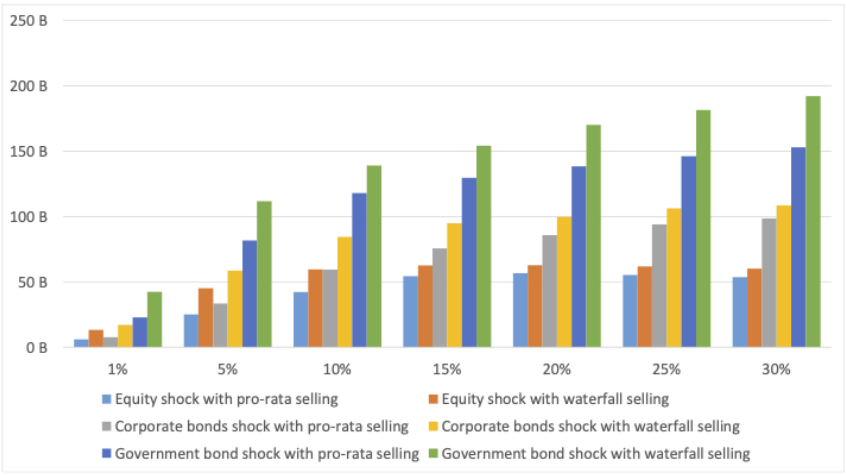


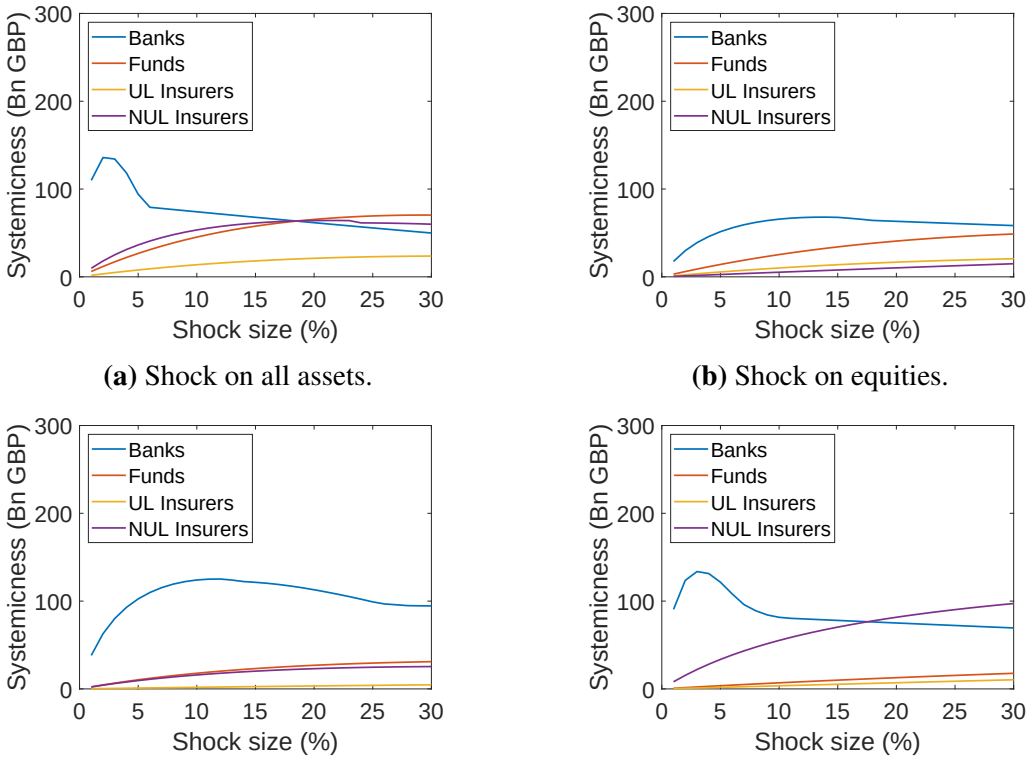
Figure 6.11: Indirect vulnerability of funds for different types of idiosyncratic shocks, when other sectors follow pro-rata vs. waterfall liquidation.

Additionally, some institutions (sectors) may have more liquid assets in common. Therefore, they may be more impacted when other institutions (sectors) prefer to follow the waterfall liquidation approach.

6.3.2.3 Systemicness

In addition to the indirect vulnerability, I would like to study the marginal contribution of each sector to the aggregate fire sale losses. To this end, I assume that there is only one sector that would actively liquidate their assets. I then compute the aggregate fire sale losses of the whole network generated solely from this sector's liquidation. In Figure 6.12, I present the systemicness of each sector for the pro-rata

liquidation across different types of idiosyncratic shock.[8]



(a) Shock on all assets.

(b) Shock on equities.

(c) Shock on corporate bonds.

(d) Shock on government bonds.

Figure 6.12: Systemicness of different sectors across different types of idiosyncratic shocks. Results are for the case of the pro-rata liquidation.

In general, I see that banks become the most systemic sector. This is not surprising, especially due to the fact that banks have the strongest constraint as was previously discussed in subsection 6.3.2. However, I also observe that other sectors may overstep the systemicness of banks in some cases. For example, the systemicness of funds and banks intersect each other at around 20% shock on equities. Additionally, I see that the systemicness of non unit-linked insurers gets beyond banks at around 18% shock on government bonds. I find that these results are related to the inverted u-shaped plots in Figure 6.8 that was previously discussed in subsection 6.3.2. In particular, as the shock gets larger, banks would only be able to liquidate less assets that remains in their balance sheet. This then results in a smaller generated price impact.

[8]The findings for the waterfall liquidation case is similar. See Appendix.

6.3.2.4 The most important institutions

From a regulatory perspective, it is useful to define the most important institutions (in term of systemicness and vulnerability) in the network. In Figure 6.13, I present a scatter plot of the importance of each institution (normalized by the total asset holdings of all institutions: £2.04 tr). I note that the most important institutions will lie in the top right corner of the plot, as these will have high vulnerability and high systemicness.

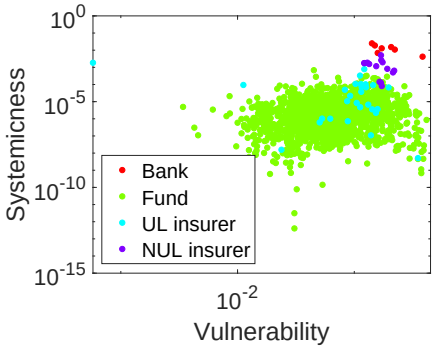


Figure 6.13: The systemicness vs. indirect vulnerability of each institution for the case of 5% initial shock on all assets. The systemicness is normalized by the total holdings of the whole network: £2.04 tr, while the indirect vulnerability is normalized by the total holdings of the corresponding institution. Plot is shown in logarithmic scale and different colours corresponds to different sectors.

As shown in Figure 6.13, institutions that belong to the same sector tend to form a cluster. In particular, I observe that the most systemically important institutions are mostly banks, followed by non unit-linked insurers. Meanwhile, funds and unit-linked insurers are relatively less important. I note that here I focus on the importance of the sector at the level of individual institutions. In fact, as was previously shown in Figure 6.12, funds may become more systemic in aggregate compared to non unit-linked insurers and even banks, because they are many.

6.3.3 Inferring the results from the network measures

I have previously discussed the results of stress simulation analysis on the UK financial system. In the following, I study the causal factors of individual institutions' contribution to the indirect vulnerability. The main scope of this analysis is to de-

termine whether the common asset holding characteristics are a useful indicator to explain the vulnerability of an institutions within my sample. In particular, I look at the following regressions

$$y_i = a + b \times X_i + \varepsilon_i, \tag{6.13}$$

where y_i is the indirect vulnerability of institution i computed using my model (see Equation 6.11), X is the set of explanatory variables, and b is the corresponding parameter vector. I consider the simple network measures (such as log(total holdings) and total links) and the portfolio similarity of each institution as the explanatory variables. I also add a dummy variable that takes on the value of 1 when the institution is a fund (and 0 otherwise). The reason of this special treatment for funds is illustrated in Figure 6.14. The figure shows the scatter plot of the total links against the (log) total holdings of each institution. In the case of banks and (unit-linked and non unit-linked) insurance companies, I observe a positive relationship between the two variables, suggesting that larger institutions tend to be more diversified. Meanwhile, in the case of funds, I find that most observations cluster around a vertical line, indicating that funds tend to be more concentrated, irrespective of their size. This is probably due to the fact that funds may often have restrictions in relation to the investment mandate. For this reason, I decided to include a dummy variable *fund* in each regression.

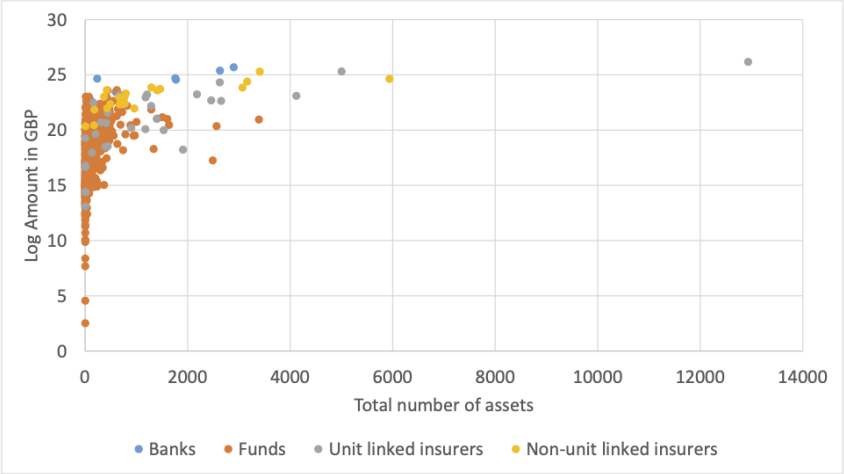


Figure 6.14: Scatter plot of the total links against the (log) total holdings of each institution.

I show the regression results for the BoE and FRB CCAR stress scenario in

Table 6.14 and Table 6.15. In Panel A, I only take the simple network measures into account. Meanwhile, in Panel B, I also consider the portfolio similarity measure. Firstly, I show from both tables that the R^2 values in Panel B are higher than those in Panel A, suggesting that the portfolio similarity measure is useful to explain more variability in the indirect vulnerability. Secondly, I find that log(total holdings) and MeanWeightedSimilarity have negative coefficients overall, suggesting that larger and more diversified institutions will become less (indirectly) vulnerable. This finding is consistent with the result in Greenwood et al. (2015) and Fricke and Fricke (2020). Note that the size of log(total holdings) is significant only for the waterfall liquidation in the FRB CCAR scenario, suggesting that some institutions with small portfolios may be more vulnerable when they have concentrated their investments mainly on the liquid assets (the first assets to be liquidated in the waterfall liquidation). Furthermore, the variable total links have also negative coefficients, suggesting that institutions with more assets in their portfolio are less vulnerable. Finally, my dummy variable *fund* confirms an additional and more significant vulnerability for the funds due to their lack of diversification.

Explanatory variables	Panel A		Panel B	
	Pro-rata	Waterfall	Pro-rata	Waterfall
Simple network measures				
log(total holdings)	$7.47e-05$	-0.000497	0.000159	-0.000423
	(0.000159)	(0.000307)	(0.000156)	(0.000306)
total links	$-1.45e-06^{*}$	$-5.78e-06^{***}$	$-1.46e-06*$	$-5.79e-06***$
	$(8.54e-07)$	$(1.64e-06)$	$(8.34e-07)$	$(1.64e-06)$
fund	-0.0144^{***}	-0.0276^{***}	$-0.0105***$	$-0.0242***$
	(0.00221)	(0.00425)	(0.00220)	(0.00431)
Portfolio similarity				
MeanWeightedSimilarity			$-0.0157***$	$-0.0138***$
			(0.00163)	(0.00319)
constant	0.0407^{***}	0.0757^{***}	$0.0397***$	$0.0748***$
	(0.00390)	(0.00750)	(0.00381)	(0.00747)
R-squared	0.025	0.022	0.070	0.032
Observations	1,923	1,923	1,923	1,923

Table 6.14: The determinants of institution-specific indirect vulnerability (normalized by the total holdings of each institution). Results are shown for the Bank of England stress scenario, and for both pro-rata and waterfall liquidation approach.

Explanatory variables	Panel A		Panel B	
	Pro-rata	Waterfall	Pro-rata	Waterfall
Simple network measures				
log(total holdings)	−0.000750	−0.00259∗∗∗	−0.000493	−0.00240∗∗∗
	(0.000599)	(0.000626)	(0.000591)	(0.000622)
total links	−5.56e−06∗	−1.06e−05∗∗∗	−5.58e−06∗	−1.06e−05∗∗∗
	(3.21e−06)	(3.35e−06)	(3.16e−06)	(3.33e−06)
fund	−0.0566∗∗∗	−0.0666∗∗∗	−0.0446∗∗∗	−0.0575∗∗∗
	(0.00829)	(0.00867)	(0.00831)	(0.00876)
Portfolio similarity				
MeanWeightedSimilarity			−0.0482∗∗∗	−0.0362∗∗∗
			(0.00616)	(0.00649)
constant	0.125∗∗∗	0.183∗∗∗	0.122∗∗∗	0.181∗∗∗
	(0.0146)	(0.0153)	(0.0144)	(0.0152)
R-squared	0.025	0.035	0.055	0.050
Observations	1,923	1,923	1,923	1,923

Table 6.15: The determinants of institution-specific indirect vulnerability (normalized by the total holdings of each institution). Results are shown for the Federal Reserve Board CCAR stress scenario, and for both pro-rata and waterfall liquidation approach.

6.4 Conclusion

In this chapter, I model the indirect contagion across UK banks and non-banks via fire sales of common asset holdings. My datasets consist of equity and bond portfolios of banks, funds and (both unit-linked and non unit-linked) insurance companies at instrument level. To this end, I assume that each financial sector may be forced to liquidate (parts of) their assets in response to losses incurred in their balance sheets. In particular, banks and non unit-linked insurers are subject to some regulatory constraints, while funds and unit-linked insurers are obliged to meet investor redemptions. Overall, the findings of this chapter contribute to a better understanding of the extent to which common asset holdings across different financial sectors become the source of financial instability.

Firstly, I find the importance of considering multiple financial sectors in the analysis. In particular, I show that ignoring the commonly held assets between banking and non-banking sector may lead to an underestimation of systemic risk by 47% in average.

Secondly, I look at the stress simulation results of the UK financial system in terms of aggregate fire sale losses, indirect vulnerability and systemicness. I con-

duct the simulation under different scenarios of initial shock and liquidation strategies. I find that the results are highly determined by the constraint and the portfolio composition of each sector. For example, banks become the most systemic in general, mainly because they liquidate a larger amount of assets relatively compared to other sectors. Moreover, I show that the yielded aggregate losses are always higher if the institutions choose to maintain their portfolio weight when liquidating their assets (pro-rata liquidation). However, I also show that the institution (sector) may become more vulnerable if other institutions (sectors) prefer to sell their most liquid assets first (waterfall liquidation). Finally, I present the map of the most systemic and the most vulnerable institutions in the UK.

Thirdly, I explore the effectiveness of network measures to explain the results of different stress simulations. Overall, I find that the portfolio similarity measure is useful to explain more variability in the indirect vulnerability. Moreover, I find that larger and more diversified institutions will become less (indirectly) vulnerable.

My findings suggest several interesting avenues for future research. First, it is important to perform similar analysis on other datasets for different countries and/or different time periods. Additionally, it is useful to incorporate more sectors into the analysis. For example, incorporating European funds that may hold similar assets to UK funds. Another important future extension is to improve the market depth calibration, such as by taking the correlations of asset prices into account.

Chapter 7

General Conclusions

In this thesis, I discussed the impact of different modelling choices on the accuracy of stress tests. I focused my study on indirect contagion channel due to common asset holdings (overlapping portfolios) between financial institutions. In particular, I studied how reliable the outcome of a stress test is, if three aspects of the test are not exactly understood. First, *the network of financial linkages* between financial institutions is often lacking, and one has to resort to network reconstruction methods to infer the network from partial information. Second, *the propagation of shocks* between financial institutions is usually modelled by means of effective dynamics, which are only approximations of the true dynamics. Third, it is mandatory to define the perimeter of the stress test, or *the types of institutions* that are included in the framework.

With regard to the first aspect, I conducted a horse race of network reconstruction methods in terms of their ability to reproduce the actual credit networks and the observed levels of systemic risk. I found that the ”best” network reconstruction method depends on the assumed criterion of interest, but the methods which preserve the actual degree distribution overall consistently perform best. I also found that the actual credit networks display the highest levels of systemic risk in many instances. Future works should study whether there are other reconstruction methods that are able to replicate the actual systemic risk levels more closely. In this thesis, I only included a small number of popular reconstruction methods, but other methods may work better.

Concerning the second aspect, I introduced a generalised stress testing model that captures a wide range of behavioural assumptions with regard to banks' liquidation dynamics under stress. The literature has proposed alternative behavioural assumptions in this regard, all of which are covered by my model. I then tested the capability of my model to predict actual bank defaults (and non-defaults). Overall, my analysis showed that the generalised model has predictive power superior to alternative benchmarks that do not account for the network of common asset holdings. I then showed how the liquidation dynamics yielding the best performing model depends on the size of the initial shock and on secondary market liquidity. I also identified, for different liquidation dynamics, the optimal number of liquidation rounds. An important follow-up question is whether individual banks follow different liquidation strategies. This might further increase the predictive accuracy of the proposed stress testing model.

Finally, with respect to the third aspect, I showed that banks and non-bank financial intermediaries have common asset holdings in their portfolios. I then looked at systemic risk arising from indirect contagion between them. I found that performing stress tests that accounts for common asset holdings between the different sectors results in fire sales losses that are 47% larger than those obtained by summing the results of sector-specific stress tests. In this thesis, I focused on different financial sectors that are located in the UK. Future research should explore the impact of incorporating also financial sectors across countries.

In summary, I showed that even if the financial network is not known, reconstruction methods do a reasonable job although they may underestimate systemic risk. Moreover, I showed that I can identify regions in the parameter space where different contagion dynamics perform better. At the same time, my analysis also showed that the model has predictive power irrespective of the assumed liquidation dynamics. Finally, I showed that neglecting cross-sector links can lead to a large underestimation of systemic risk.

Appendix A

On Network Reconstruction

A.1 Weight Allocation Methods

I use RAS (Blien and Graef (1998)) method to distribute the observed credit volumes across links for the generated adjacency matrix of CM1 and CM2. Previously, I experimented with different weight allocation approaches defined below and finally find that RAS generally performed best in my analysis (in term of corresponding L_1-error).

A.2 Robustness Checks: Systemic Risk Analysis on Other Models

For the purpose of finding out how the systemic risk analysis might vary if leverage targeting model (as in Greenwood et al. (2015)) and leverage targeting with threshold model (as in Cont and Schaanning (2017)) are used, I also performed the same exercise with these other models. I find that the rank ordering of the different methods are generally consistent with those presented in the main text.

Weight allocation method	Definition
RAS (Blien and Graef (1998))	Column constraint $$\hat{w}_{ij}(t+1) = \frac{\hat{w}_{ij}(t)}{\hat{s}(t)^B_i,} \times s^B_i,$$ Row constraint $$\hat{w}_{ij}(t+1) = \frac{\hat{w}_{ij}(t)}{\hat{s}(t)^F_i,} \times s^F_i,$$ where t is the respective iteration step.
Linear Programming (Mohr and Polenske (1987))	$$Maximize \sum_{i=1}^{N}\sum_{j=1}^{M} c_{ij}\hat{w}_{ij}$$ $$subject\,to \sum_{i=1}^{N} \hat{w}_{ij} = s^F_j \qquad (j = 1,\ldots,n)$$ $$\sum_{j=1}^{M} \hat{w}_{ij} = s^B_i \qquad [i = 1,\ldots,(m-1)]$$ $$\hat{w}_{ij} > (c_{ij})(\varepsilon)$$ where $\bar{w}_{ij} > 0 \rightarrow c_{ij} = 1$, $\bar{w}_{ij} = 0 \rightarrow c_{ij} = 0$
Maximum Flow (Cormen et al. (2009))	See Gandy and Veraart (2016) for the discussion on how to transform this into a maximum flow problem.

Table A.1: Summary of different weight allocation methods for the bank-firm network. Note that I can define the same methods for the bank-industry network.

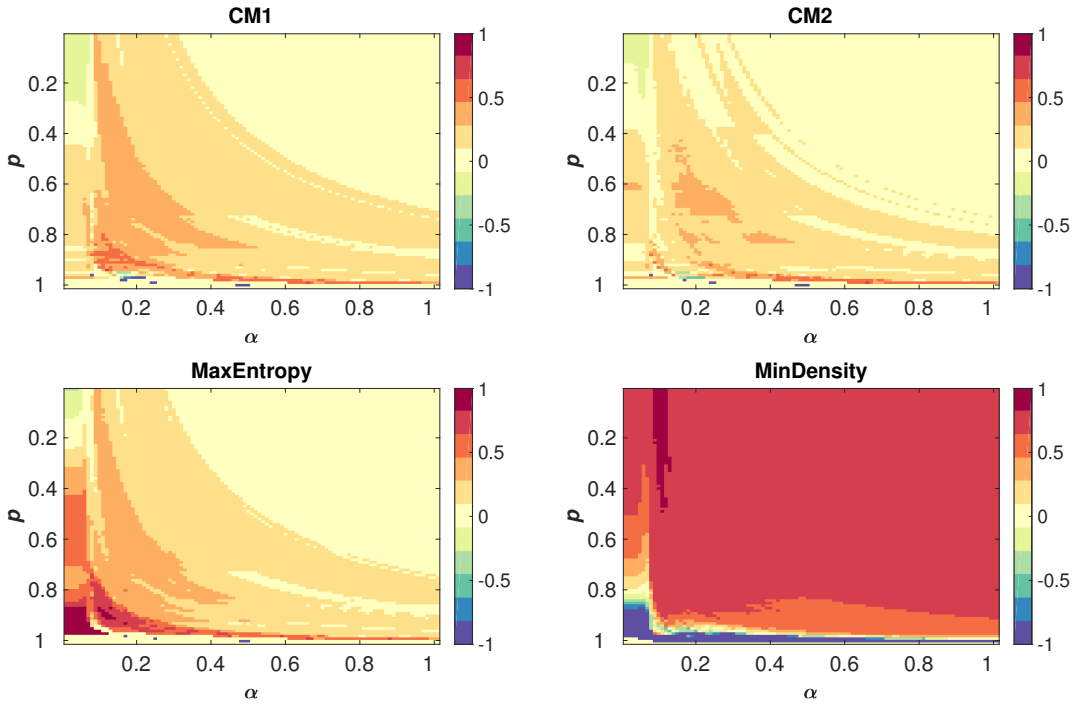


Figure A.1: (
small to large market impact) and $p \in [0,1]$ (large to small initial shock). Leverage targeting model Greenwood et al. (2015) is used.]Relative difference of the probability of default between actual network and the null models (D_r) at the aggregated level for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Leverage targeting model Greenwood et al. (2015) is used. Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation.

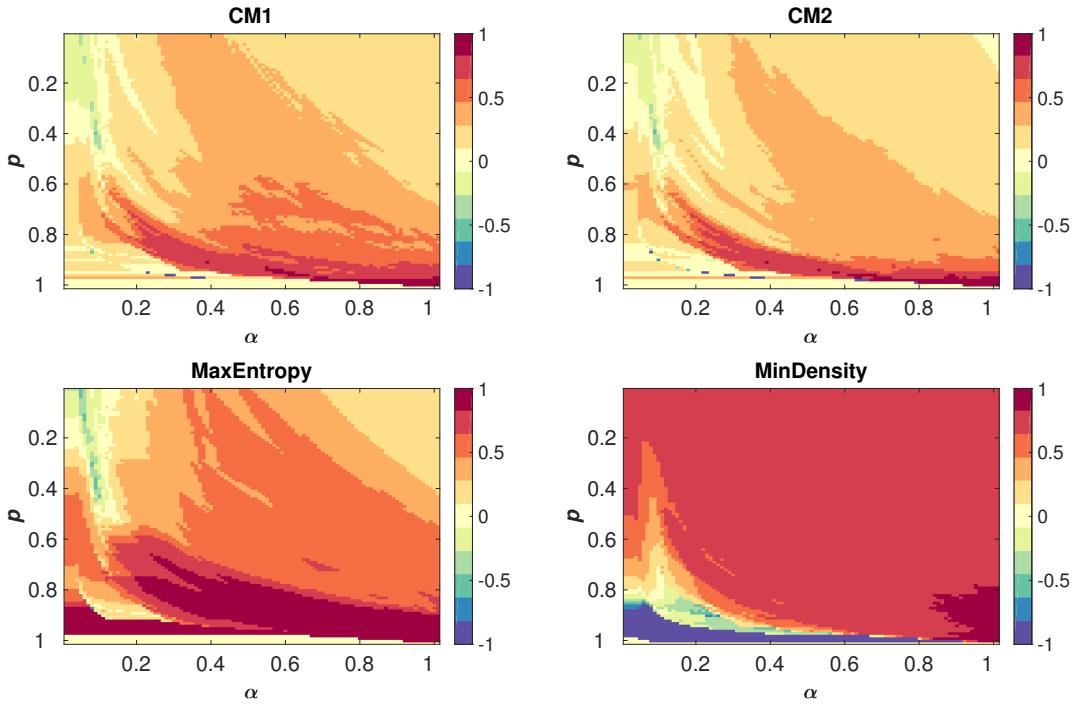


Figure A.2: (
small to large market impact) and $p \in [0,1]$ (large to small initial shock). Threshold model (Cont and Schaanning (2017)) is used.]Relative difference of the probability of default between actual network and the null models (D_r) at the aggregated level for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Threshold model (Cont and Schaanning (2017)) is used. Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation.

A.3 Additional Results: All Values of Parameters

In Table 4.7, I previously computed the average of P_d across a range of restricted values of parameters that I defined as $p \in [0.6, 1]$ and $\alpha \in [0.6, 1]$. In the following, I again compute the average of P_d but for all possible values of parameters (including those that I define as restricted and not restricted).

Rank	Disaggregated		Aggregated		Intermediate	
	Null model	$\overline{P_d}$	**Null model**	$\overline{P_d}$	**Null model**	$\overline{P_d}$
1	Actual	0.393 (0.254)	Actual	0.360 (0.230)	Actual	0.360 (0.230)
2	CM1	0.301 (0.202)	CM1	0.218 (0.156)	MinDensity	0.358 (0.217)
3	CM2	0.243 (0.176)	CM2	0.217 (0.157)	CM1	0.275 (0.182)
4	MaxEntropy	0.190 (0.149)	MinDensity	0.202 (0.122)	CM2	0.241 (0.174)
5	MinDensity	0.140 (0.096)	MaxEntropy	0.190 (0.149)	MaxEntropy	0.190 (0.149)

Table A.1: Rank of the actual networks and the corresponding null models at different aggregation levels as in Table 4.7 for the 2010 data. However, as opposed to Table 4.7, where I consider only the restricted values of parameters, here I calculate the average of P_d across all possible parameter combinations: $p \in \{0, 0.01, 0.02, \ldots, 1\}$ and $\alpha \in \{0, 0.01, 0.02, \ldots, 1\}$. Rank 1 corresponds to the most risky network. $\overline{P_d}$ denotes the average, and the value inside the bracket is its corresponding standard deviation.

A.4 Additional Results: Wilcoxon Signed Rank Test on the Networks

I formally test whether the difference between each network P_d is significant by running a two-sided Wilcoxon signed rank test on each pair of the actual network and the null model. In the tables below, I show the corresponding p-values of each test for different range of p and α. In Table A.1, I consider a range of parameters **within** the restricted values that I defined in the main text ($p \in [0.6, 0.8]$ and $\alpha \in [0.6, 0.8]$). Meanwhile, in Table A.2, I consider a range of parameters **outside** the restricted values that I previously defined. In particular, I consider all possible values of the initial shock ($p \in [0, 1]$) and relatively liquid assets ($\alpha \in [0, 0.5]$). Overall, I find that the difference between each network P_d is significant, except for CM1 and CM2 in some instances.

Disaggregated	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Aggregated	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		⋄0.389	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Intermediate	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Table A.1: P-value of a two-sided Wilcoxon signed rank test on each pair of the network. A sufficiently small p-value indicates that the test rejects the null hypothesis that the difference between the pairs follow a symmetric distribution around zero, thus the two networks have significantly different P_ds. Meanwhile, a large p-value indicates that the test fails to reject the null hypothesis, thus the difference between the two networks P_ds is not significant. Here I test the P_d value of each network for $p \in [0.6, 0.8]$ and $\alpha \in [0.6, 0.8]$. I highlight the p-value above 0.05 using the $\diamond$ symbol.

Disaggregated	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Aggregated	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		$\diamond$0.158	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Intermediate	CM1	CM2	MaxEn-tropy	MinDen-sity
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Table A.2: *P*-value of a two-sided Wilcoxon signed rank test on each pair of the network as in Table A.1. However, here I consider $p \in [0, 1]$ and $\alpha \in [0, 0.5]$.

Appendix B

On Dynamics of Contagion

B.1 Logistic Regression

In chapter 5, I plot the ROC curve of a standard logistic regression model. The idea of this model is that I am interested in the conditional probability of observing bank i's default, which I denote as $Prob(y_i = 1|X = x_i)$, where x_i is a vector of leverage and total assets of bank i (log transformed) in my case. Logistic regression specifies this probability as:

$$Prob(y_i = 1|X = x_i) = \frac{1}{1 + \exp^{(-x_i\beta)}}, \tag{B.1}$$

where β is a vector of parameter estimates for the explanatory variables (estimated via maximum likelihood).

I present a summary of the results of a logistic regression model in Table B.1. As it can be seen from the table, the explanatory variables (leverage and total assets) are significantly positively associated with the probability of default.

B.2 Enhanced logistic regression

Table B.2 shows the results of the enhanced logistic regression model with additional explanatory variables.

	Coefficient	**Standard error**	***t*-stat**
Intercept	-8.74	0.60	-14.59 ***
log(Leverage)	1.15	0.20	5.77 ***
log(TotalAssets)	0.24	0.04	6.70 ***
Pseudo-R^2	0.01		
Obs.	7,783		

Table B.1: Results from a logistic regression model to explain bank failures where the dependent variable takes on a value of one if a bank failed, and a value of zero otherwise. The explanatory variables are leverage and total assets (log transformed). *** indicates statistical significance at the 1% level.

	Coefficient	**Standard error**	***t*-stat**
Intercept	-7.28	2.94	-2.47^{*}
Asset-1	7.83	3.02	2.59^{**}
Asset-2	-2.04	3.74	-0.55
Asset-3	-2.33	3.07	-0.76
Asset-4	5.35	3.14	1.70
Asset-5	0.71	3.02	0.24
Asset-6	-0.11	3.33	-0.03
Asset-7	-0.23	3.06	-0.08
Asset-8	-10.21	4.48	-2.28^{*}
Asset-9	1.56	5.05	0.31
Asset-11	-2.45	3.28	-0.75
Asset-12	-1.75	3.04	-0.58
Asset-14	-6.68	4.34	-1.54
log(Leverage)	0.02	0.06	0.41
log(TotalAssets)	1.64	0.24	6.76^{***}
Pseudo-R^2	0.17		
Obs.	7,783		

Table B.2: Results from the enhanced logistic regression model. The dependent variable takes a value of one if a bank failed, and zero otherwise. The explanatory variables are leverage, total assets (both log-transformed) and the portfolio share of each asset class (excluding asset classes 10 and 13). *, ** and *** indicates statistical significance at the 5%, 1% and 0.1% level.

B.3 Initial shock versus contagion dynamics

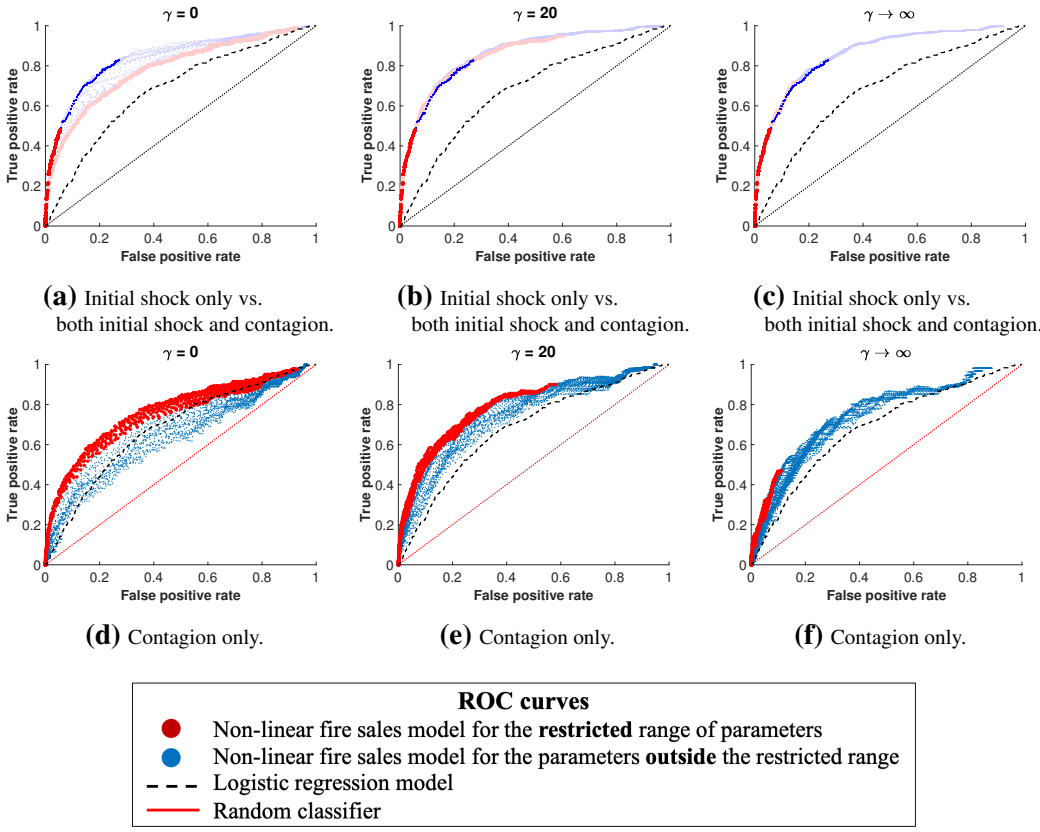
In chapter 5, I focus on bank failures identified from the contagion dynamics alone, and ignored failures due to the initial shock. In Figure B.1, I show the comparison between defaults due to three different categories: (i) initial shock only, (ii) both initial shock and contagion, and (iii) contagion only. First, I find that the results of the second category are superior compared to those I presented in the main text. In other words, the results shown in the main text likely underestimate the performance of the network approach.

Second, including the contagion mechanism improves the model output by identifying defaults that are not due to the initial shock. Recall that values closer to the top right corner of the TPR/FPR space correspond to a larger number of banks predicted to default. For the case of the initial shock only, the dots cover only a limited range within the TPR/FPR space. Meanwhile, for the case of both initial shock and contagion, the dots cover a much broader range.

Moreover, Figure B.2 shows the total losses of all banks due to initial shock versus due to contagion only, for $\gamma = 0$ (leverage targeting model). The results indicate that the latter is relatively larger, indicating the importance of contagion dynamics in explaining the results.

B.4 Shocking Less Relevant Asset Classes

I found that two asset classes (1 and 5) were most relevant in terms of my model application. For the sake of completeness, here I show results when imposing an initial shock on asset 8 (loans to individuals). Figure B.3 shows that all dots in the ROC space now lie close to the diagonal or even below it. In other words, when imposing an initial shock on less relevant assets, the model performs very similar to or worse than a random classifier.



(a) Initial shock only vs. both initial shock and contagion.

(b) Initial shock only vs. both initial shock and contagion.

(c) Initial shock only vs. both initial shock and contagion.

(d) Contagion only.

(e) Contagion only.

(f) Contagion only.

Figure B.1: ROC curves of the defaulted banks predicted by the model with $\gamma = 0$ (left panels), $\gamma = 20$ (middle panels) and $\gamma \to \infty$ (right panels), with an initial shock on asset class 1 (loans for construction and land development). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. (a)(b)(c) ROC curves of bank failures identified due to the initial shock only vs. initial shock and contagion (darker color vs lighter color), (d)(e)(f) from contagion only. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate.

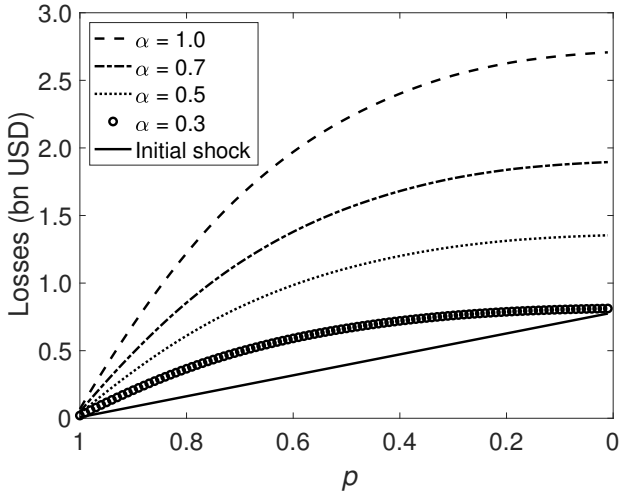


Figure B.2: Total losses (in billion USD) due to the initial shock and due to the contagion dynamics only (first round of liquidation), for $\gamma = 0$. Results for initial shock on asset class 1 (loans for construction and land development). The total pre-shock assets in the network are 12.7 bn USD. Parameter p is the post-shock value of the asset (as a fraction of the pre-shock value), where a lower p corresponds to a larger initial shock. α is the market impact parameter, where a higher α corresponds to a more illiquid asset.

B.5 Multiple liquidation rounds outside the restricted parameter range

In the main text, I present the results for $p = \alpha = 0.7$ that are within the restricted range of parameters. In the following, I show the corresponding results for $p = \alpha = 0.5$ that are outside the restricted range of parameters. I find that the results are consistent to those in the main text.

B.6 Model Accuracy with a Non-Linear Market Impact Function

In the main text, I consider a linear market impact function (Ψ_j). Here I look at the results of non-linear Ψ_j. To this end, I use a similar functional form as in Caccioli et al. (2014), that is:

$$\Psi_j(x_j^t) = e^{-\alpha x_j^t},$$

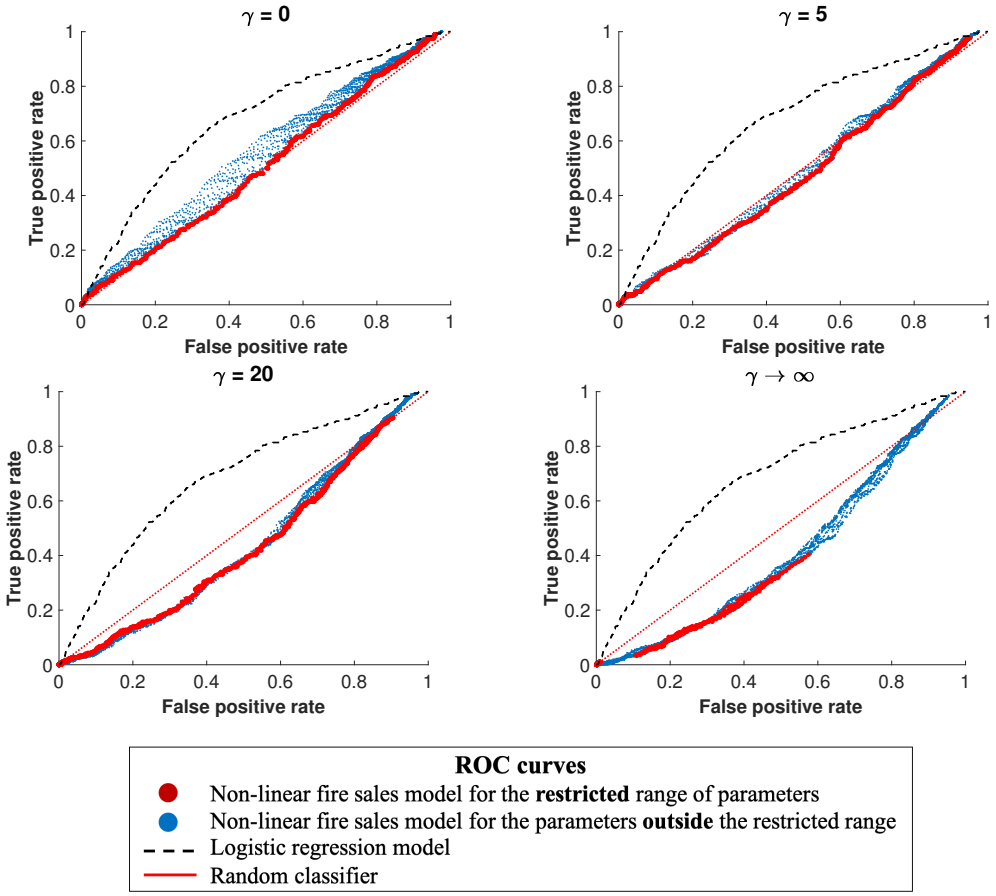


Figure B.3: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with an initial shock on loans to individuals (asset class 8). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

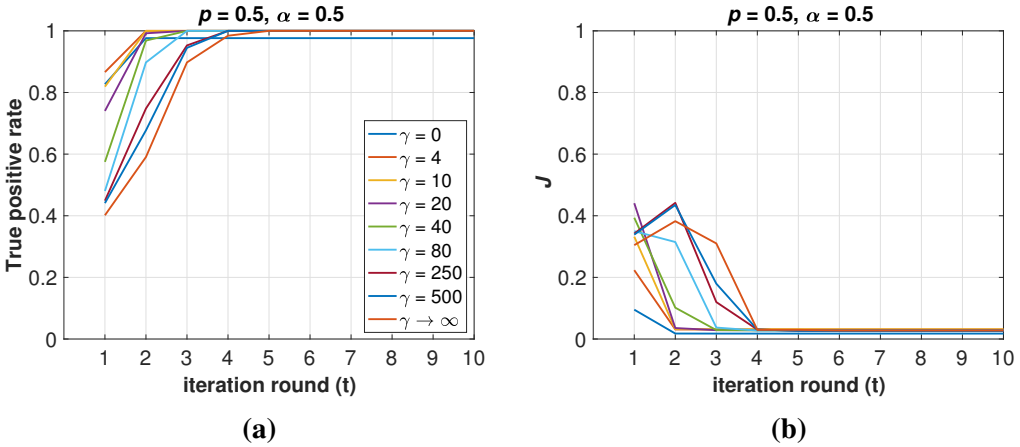


Figure B.4: (a) The number of bank failures that is correctly identified by the model (normalized to the actual number of bank failures), as a function of γ and iteration round (t). (b) The corresponding value of J as a function of γ and iteration round (t). The results correspond to a combination of p and α in the non-restricted parameter range ($p = 0.5$ and $\alpha = 0.5$).

for a range of α as in the main text, $\alpha \in \{0, 0.01, 0.02, ..., 0.99, 1\}$. I plot the corresponding ROC curves of this choice of Ψ_j in Figure B.5 (for initial shock on asset: loans for construction and land development) and in Figure B.6 (for initial shock on asset: loans secured by non-farm non-residential properties). I find that the results of linear and non-linear market impact are qualitatively similar (see Figure 5.4 and Figure 5.5 to compare with linear market impact).

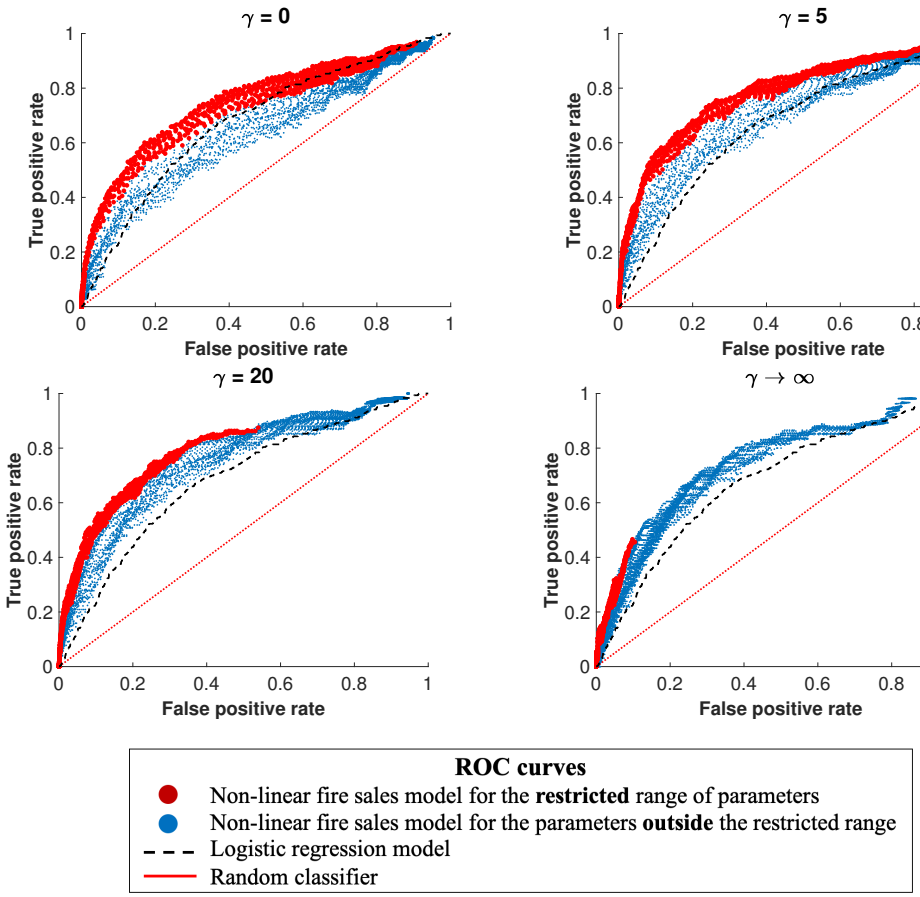


Figure B.5: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with an initial shock on loans for construction and land development (asset class 1). The plot is similar to Figure 5.4, but here I consider a non-linear market impact function. Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

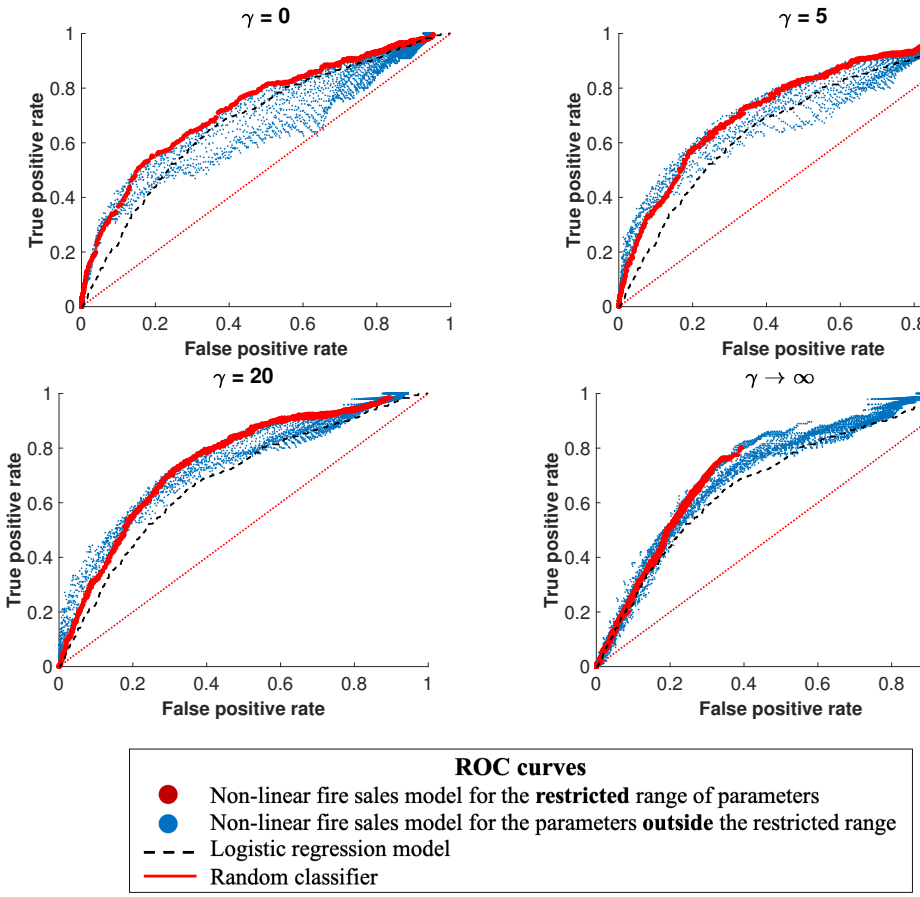


Figure B.6: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \to \infty$, resp., with an initial shock on loans secured by non-farm non-residential properties (asset class 5). The plot is similar to Figure 5.5, but here I consider a non-linear market impact function. Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). I highlight the results for the restricted range of parameters (low initial shock and high market impact) in red; blue corresponds to parameter combinations outside this range. The black dashed line is the ROC curve of a corresponding logistic regression model with bank leverage and total assets (log-transformed) as explanatory variables, and the red diagonal line is the ROC curve of a random classifier. A model closer to the top left corner of the TPR/FPR space is considered more accurate. Here I consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

Appendix C

On System-Wide Stress Tests

C.1 Indirect vulnerability of unit-linked and non unit-linked insurers

In the main text, I have discussed the indirect vulnerability resulted from the pro-rata vs. waterfall liquidation approach, for the case of banks and funds. For what follows, I present the similar results, for the case of unit-linked and non unit-linked insurance companies.

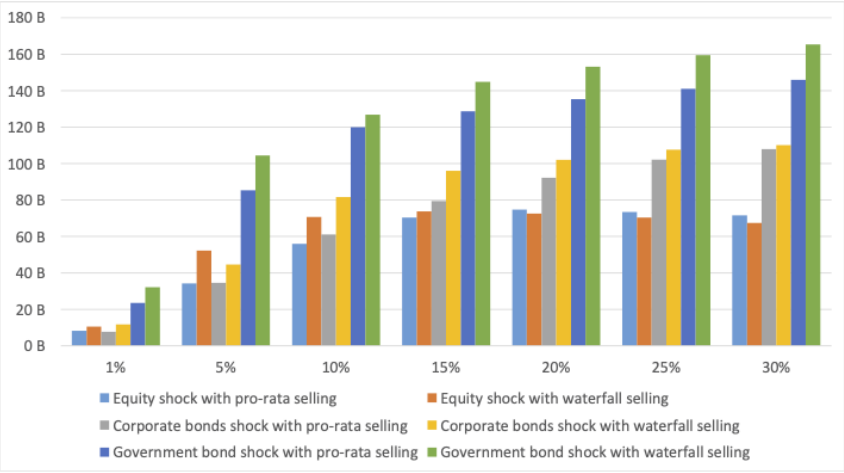


Figure C.1: Indirect vulnerability of unit-linked insurers for different types of idiosyncratic shocks, when other sectors follow pro-rata vs. waterfall liquidation.

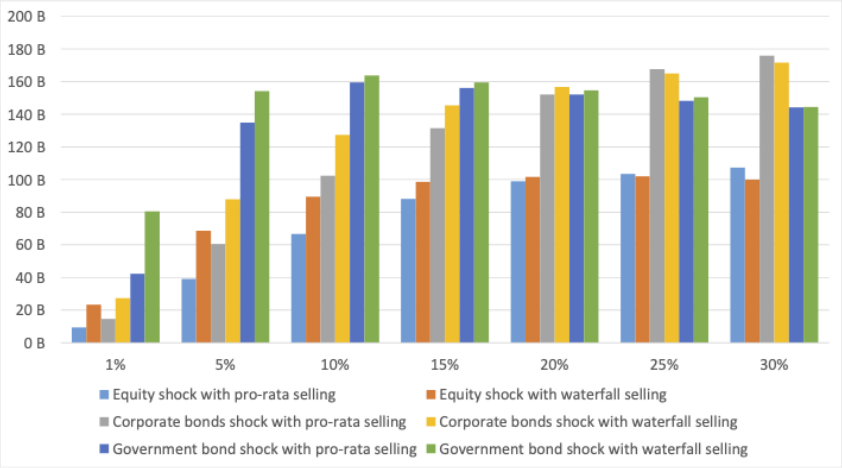
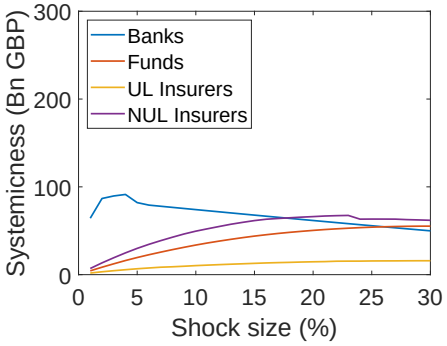


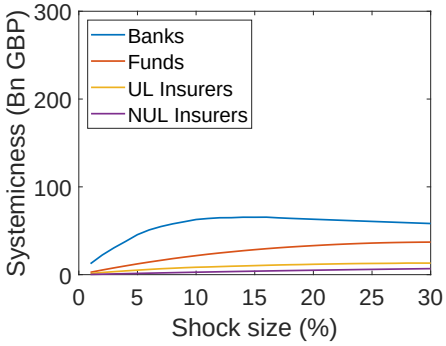
Figure C.2: Indirect vulnerability of non unit-linked insurers for different types of idiosyncratic shocks, when other sectors follow pro-rata vs. waterfall liquidation.

C.2 Systemicness

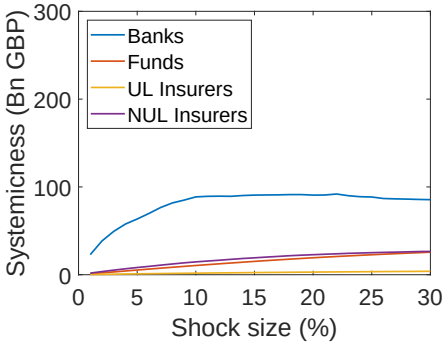
In the main text, I have discussed the systemicness of different sector across different types of idiosyncratic shocks, for the case of the pro-rata liquidation approach. For what follows, I discuss the similar results, for the case of the waterfall liquidation.



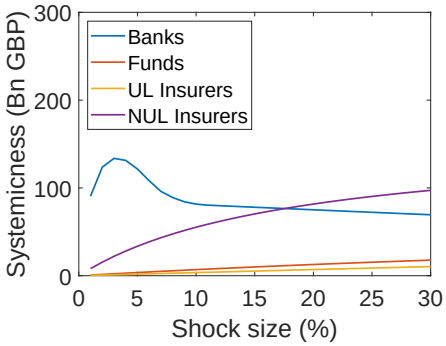
(a) Shock on all assets.



(b) Shock on equities.



(c) Shock on corporate bonds.



(d) Shock on government bonds.

Figure C.3: Systemicness of different sectors across different types of idiosyncratic shocks. Results are for the case of waterfall liquidation.

Bibliography

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437.

Aikman, D., Chichkanov, P., Douglas, G., Georgiev, Y., Howat, J., and King, B. (2019). System-wide stress simulation. *Bank of England Staff Working Paper No. 809*.

Allen, F. and Gale, D. (1999). The asian crisis and the process of financial contagion. *Journal of Financial Regulation and Compliance*, 7(3).

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239.

Anand, K., Craig, B., von Peter, G., and Von Peter, G. (2015). Filling in the blanks: network structure and interbank contagion. *Quantitative Finance*, 15(4):625–636.

Anand, K., Van Lelyveld, I., Banai, Á., Friedrich, S., Garratt, R., Hałaj, G., Fique, J., Hansen, I., Jaramillo, S. M., Lee, H., Molina-Borboa, J. L., Nobili, S., Rajan, S., Salakhova, D., Silva, T. C., Silvestri, L., and De Souza, S. R. S. (2018). The missing links: A global study on uncovering financial network structures from partial data. *Journal of Financial Stability*, 35(April 2018):107–119.

Anderson, R. W. (2016). Stress testing and macroprudential regulation: A transatlantic assessment. In Anderson, R. W., editor, *Stress Testing and Macroprudential Regulation: A Transatlantic Assessment*, pages 1–29. CEPR Press, London.

Anderson, R. W., Danielsson, J., Baba, C., Das, U. S., Kang, H., and Basurto, M. A. S. (2018). Macroprudential stress tests and policies: searching for robust and implementable frameworks. *IMF Working Paper*, WP/18/197.

Aymanns, C., Farmer, J. D., Kleinnijenhuis, A. M., and Wetzer, T. (2018). Models of financial stability and their application in stress tests. *Handbook of Computational Economics*, 4:329–391.

Bank of England (2016). Financial stability report.

Banwo, O., Caccioli, F., Harrald, P., and Medda, F. (2016). The effect of heterogeneity on financial contagion due to overlapping portfolios. *Advances in Complex Systems*, 19(08):1650016.

Baranova, Y., Coen, J., Lowe, P., Noss, J., and Silvestri, L. (2017). Simulating stress across the financial system: the resilience of corporate bond markets and the role of investment funds. *Bank of England Financial Stability Paper No. 42*.

Baranova, Y., Douglas, G., and Silvestri, L. (2019). Simulating stress in the uk corporate bond market: investor behaviour and asset fire-sales. *Bank of England Staff Working Paper No. 803*.

Bardoscia, M., Caccioli, F., Perotti, J. I., Vivaldo, G., and Caldarelli, G. (2016). Distress propagation in complex networks: the case of non-linear debtrank. *PLoS ONE*, 11(10).

Barucca, P., Mahmood, T., and Silvestri, L. (2020). Common asset holdings and systemic vulnerability across multiple types of financial institution. *Journal of Financial Stability*, Forthcoming.

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., and Scheffer, M. (2016). Complexity theory and financial regulation. *Science*, 351(6275):818–819.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012). DebtRank: too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2(541).

Baudino, P., Goetschmann, R., Henry, J., Taniguchi, K., and Zhu, W. (2018). Stress-testing banks - a comparative analysis. *FSI Insights on policy implementation No 12*, pages 1–36.

BIS (2015). Making supervisory stress tests more macroprudential: considering liquidity and solvency interactions and systemic risk. *BCBS Working Papers No 29*.

Bisias, D., Flood, M., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. *Office of Financial Research Working Paper*.

Blien, U. and Graef, F. (1998). Entropy optimizing methods for the estimation of tables. In *Classification, Data Analysis, and Data Highways*, pages 3–15.

Bryant, J. (1980). A model of reserves, bank runs, and deposit insurance. *Journal of Banking and Finance*, 4:335–344.

Caccioli, F., Barucca, P., and Kobayashi, T. (2018). Network models of financial systemic risk: a review. *Journal of Computational Social Science*, 1:81–114.

Caccioli, F., Catanach, T. A., and Farmer, J. D. (2012). Heterogeneity, correlations and financial contagion. *Advances in Complex Systems*, 15(supp02):1250058.

Caccioli, F., Farmer, J. D., Foti, N., and Rockmore, D. (2015). Overlapping portfolios, contagion, and financial stability. *Journal of Economic Dynamics and Control*, 51:50–63.

Caccioli, F., Shrestha, M., Moore, C., and Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, 46:233–245.

Cappiello, L., Rousova, L. F., and Montagna, M. (2015). Systemic risk, contagion and financial networks. *European Central Bank Special Feature*, pages 147–158.

Cavestany, R. and Rodríguez, D. (2015). Backtesting, stress testing and sensitivity analysis. In Cavestany, R., Boultwood, B., and Escudero, L., editors, *Operational Risk Capital Models*. Risk Books.

Cetorelli, N., Duarte, F., and Eisenbach, T. M. (2016). Are asset managers vulnerable to fire sales? *Federal Reserve Bank of New York, Liberty Street Economics.*

Chari, V. V. (1988). Banking panics, information, and rational expectations equilibrium. *Journal of Finance*, 43:749:760.

Chernenko, S. and Sunderam, A. (2016). Liquidity transformation in asset management: evidence from the cash holdings of mutual funds. *ESRB Working Paper Series No 23*.

Cifuentes, R., Ferrucci, G., and Shin, H. S. (2005). Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2-3):556–566.

Cimini, G., Squartini, T., Gabrielli, A., and Garlaschelli, D. (2015a). Estimating topological properties of weighted networks from limited information. *Physical Review E*, 92(4):1539–3755.

Cimini, G., Squartini, T., Garlaschelli, D., and Gabrielli, A. (2015b). Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports*, 5:15758.

Coen, J., Lepore, C., and Schaanning, E. (2019). Taking regulation seriously: fire sales under solvency and liquidity constraints. *Staff Working Paper No. 793*.

Cole, R. A. and White, L. J. (2012). Déjà vu all over again: the causes of U.S. commercial bank failures this time around. *Journal of Financial Services Research*, 42(1-2):5–29.

Cont, R. and Schaanning, E. (2017). Fire sales, indirect contagion and systemic stress testing. *Norges Bank Working Paper*, 2.

Cont, R. and Wagalath, L. (2016). Fire sales forensics: measuring endogeneous risk. *Mathematical Finance*, 26(4):835–866.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms Third Edition*. The MIT Press.

Coval, J. and Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512.

Czech, R. and Roberts-Sklar, M. (2019). Investor behaviour and reaching for yield: evidence from the sterling corporate bond market. *Financial Markets, Institutions and Instruments*, 28(5):347–379.

Daníelsson, J. (2011). Backtesting and stress testing. In *Backtesting and Stress Testing*, pages 143–166. John Wiley & Sons Ltd, West Sussex.

De Bandt, O., Hartmann, P., and Peydró, J. L. (2012). *Systemic Risk in Banking: An Update*. Number January 2020.

De Masi, G. and Gallegati, M. (2012). Bank-firms topology in italy. *Empirical Economics*, 43:851–866.

Dees, S., Henry, J., and Martin, R. (2017). STAMP: stress-test analytics for macro-prudential purposes in the euro area. *European Central Bank*.

Dent, K., Westwood, B., and Segoviano, M. (2016). Stress testing of banks: an introduction. *Bank of England Quarterly Bulletin*, 56(3).

Di Gangi, D., Lillo, F., and Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control*, 94:117–141.

Diamond, D. W. and Dybvig, P. H. (1983). Liquidity shortages and banking crises. *Journal of Finance*, 60:615–647.

Diamond, D. W. and Rajan, R. G. (2011). Fear of fire sales, illiquidity seeking, and credit freezes. *The Quarterly Journal of Economics*, 126(2):557–591.

Douglas, G., Noss, J., and Vause, N. (2017). The impact of solvency ii regulations on life insurers' investment behaviour. *Bank of England Staff Working Paper No. 664*.

Douglas, G. and Roberts-Sklar, M. (2018). What drives uk defined benefit pension funds' investment behaviour? *Bank of England Staff Working Paper No. 757*.

Drucker, S. and Puri, M. (2009). On loan sales, loan contracting, and lending relationships. *Review of Financial Studies*, 22(7):2835–2872.

Duarte, F. and Eisenbach, T. M. (2015). Fire-sale spillovers and systemic risk. *Federal Reserve Bank of New York Staff Reports No. 645*.

Egan, J. P. (1975). *Signal Detection Theory and Roc Analysis, Series in Cognition and Perception*. Academic Press, New York.

Eisenberg, L. and Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.

Ellul, A., Jotikasthira, C., and Lundblad, C. T. (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics*, 101(3):596–620.

European Central Bank (2004). *Annual Report: 2004*.

European Central Bank (2014). Structural and systemic risk features of euro area investment funds. *Financial Stability Review*.

Farmer, J. D., Kleinnijenhuis, A. M., Nahai-Williamson, P., and Wetzer, T. (2020). Foundations of system-wide financial stress testing with heterogeneous institutions. *Bank of England Staff Working Paper No. 861*.

Fosdick, B. K., Larremore, D. B., Nishimura, J., and Ugander, J. (2018). Configuring random graph models with fixed degree sequences. *SIAM Rev.*, 60(2):315–355.

Fouque, J.-P. and Langsam, J. A. (2013). *Handbook of Systemic Risk*. Cambridge University Press, New York.

Fricke, C. and Fricke, D. (2020). Vulnerable asset management? the case of mutual funds. *Journal of Financial Stability*, Forthcoming.

Fricke, D. (2019). Are specialist funds "special"? *Financial Management*, 48(2):441–472.

Fricke, D. and Roukny, T. (2020). Generalists and specialists in the credit market. *Journal of Banking and Finance*, 112(March 2020).

Furfine, C. H. (2003). Interbank exposures: quantifying the risk of contagion. *Journal of Money, Credit and Banking*, 35(1):111–128.

Gai, P. and Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466:2401–2423.

Gale, D. and Yorulmazer, T. (2013). Liquidity hoarding. *Theoretical Economics*, 8:291–324.

Gandy, A. and Veraart, L. A. M. (2016). A bayesian methodology for systemic risk assessment in financial networks. *Management Science*, pages 1–20.

Gandy, A. and Veraart, L. A. M. (2017). Adjustable network reconstruction with applications to cds exposures.

Glasserman, P. and Young, H. P. (2015). How likely is contagion in financial networks? *Journal of Banking and Finance*, 50:383–399.

Glasserman, P. and Young, H. P. (2016). Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831.

Greenwood, R., Landier, A., and Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics*, 115(3):471–485.

Grigat, D. and Caccioli, F. (2017). Reverse stress testing interbank networks. *Scientific Reports*, 7(15616).

Gualdi, S., Cimini, G., Primicerio, K., Clemente, R. D., and Challet, D. (2016). Statistically validated network of portfolio overlaps and systemic risk. *Scientific Reports*, 6:39467.

Haldane, A. G. (2015). On microscopes and telescopes. speech available at http://www.bankofengland.co.uk.

Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469:351–355.

Hale, G., Krainer, J., and Mccarthy, E. (2015). Aggregation level in stress testing models. *FRBSF Working Paper*.

Hatzopoulos, V., Iori, G., Mantegna, R. N., Miccichè, S., and Tumminello, M. (2015). Quantifying preferential trading in the e-mid interbank market. *Quantitative Finance*, 15(4):693–710.

Heise, S. and Kühn, R. (2012). Derivatives and credit contagion in interconnected networks. *Eur. Phys. J. B*, 85(115).

Huang, X., Vodenska, I., Havlin, S., and Stanley, H. E. (2013). Cascading failures in bi-partite graphs: model for systemic risk propagation. *Scientific Reports*, 3(1219).

Hüser, A.-C. (2015). Too interconnected to fail: A survey of the interbank networks literature. *Journal of Network Theory in Finance*, 1(3):1–50.

International Monetary Fund (2012). Macrofinancial stress testing - principles and practices. Technical report.

International Monetary Fund (2015). Global financial stability review. *Technical report*.

Iori, G., Jafarey, S., and Padilla, F. G. (2006). Systemic risk on the interbank market. *Journal of Economic Behavior & Organization*, 61(4):525–542.

Iori, G., Mantegna, R. N., Marotta, L., Miccichè, S., Porter, J., and Tumminello, M. (2015). Networked relationships in the e-mid interbank market: A trading model with memory. *Journal of Economic Dynamics and Control*, 50:98–116.

Jiang, H., Li, D., and Wang, A. W. (2017). Dynamic liquidity management by corporate bond mutual funds. *Available at SSRN: https://ssrn.com/abstract=2776829*.

Kaufman, G. G. and Scott, K. E. (2003). What is systemic risk, and do bank regulators retard or contribute to it? *The Independent Review*, 7(3):371–391.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? evidence from subprime loans. *The Quarterly Journal of Economics*, 125(1):307–362.

Khandani, A. and Lo, A. (2011). What happened to the quants in august 2007? evidence from factors and transactions data. *Journal of Financial Markets*, 14(1):1–46.

Levy-Carciente, S., Kenett, D. Y., Avakian, A., Stanley, H. E., and Havlin, S. (2015). Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance*, 59:164–181.

Lillo, F. and Pirino, D. (2015). The impact of systemic and illiquidity risk on financing with risky collateral. *Journal of Economic Dynamics and Control*, 50:180–202.

Manconi, A., Massa, M., and Yasuda, A. (2012). The role of institutional investors in propagating the crisis of 2007–2008. *Journal of Financial Economics*, 104(3):491–518.

Marotta, L., Miccichè, S., Fujiwara, Y., Iyetomi, H., Aoyama, H., Gallegati, M., and Mantegna, R. N. (2015). Bank-firm credit network in japan: an analysis of a bipartite network. *PLoS ONE*, 10(5):e0123079.

Martínez-Jaramillo, S., Pé Rez Pé Rez, O., Avila Embriz, F., Ló Pez, F., and Dey, G. (2010). Systemic risk, financial contagion and financial fragility. *Journal of Economic Dynamics and Control*, 34:2358–2374.

Mastrandrea, R., Squartini, T., Fagiolo, G., and Garlaschelli, D. (2014). Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics*, 16(4):043022.

Mastromatteo, I., Zarinelli, E., and Marsili, M. (2012). Reconstruction of financial networks for robust estimation of systemic risk. *Journal of Statistical Mechanics: Theory and Experiment*, 12:1742–5468.

May, R. M., Levin, S. A., and Sugihara, G. (2008). Ecology for bankers. *Nature*, 451:893–895.

Mazzarisi, P. and Lillo, F. (2017). Methods for reconstructing interbank networks from limited information: A comparison. In Abergel, F., Aoyama, H., Chakrabarti, B. K., Chakraborti, A., Deo, N., Raina, D., and Vodenska, I., editors, *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 201–215. Springer, Cham.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127.

Mohr, M. and Polenske, K. R. (1987). A linear programming approach to solving infeasible ras problems. *Journal of Regional Science*, 27(4):587–603.

Musmeci, N., Battiston, S., Caldarelli, G., Puliga, M., and Gabrielli, A. (2013). Bootstrapping topological properties and systemic risk of complex networks using the fitness model. *J Stat Phys*, 151(3-4):720–734.

Neveu, A. R. (2016). A survey of network-based analysis and systemic risk measurement. *Journal of Economic Interaction and Coordination*.

Niepmann, F. and Stebunovs, V. (2018). Modeling your stress away. *International Finance Discussion Papers*, 1232(7).

Philippon, T., Pessarossi, P., and Camara, B. (2017). Backtesting european stress tests. *NBER Working Paper No. 23083*.

Poledna, S., Martínez-Jaramillo, S., Caccioli, F., and Thurner, S. (2018). Quantification of systemic risk from overlapping portfolios in the financial system. *arXiv:1802.0031*.

Puhr, C. and Schmitz, S. W. (2014). A view from the top: the interaction between solvency and liquidity stress. *Journal of Risk Management in Financial Institutions*, 7(1):38–51.

Pulvino, T. C. (1998). Do asset fire sales exist? an empirical investigation of commercial aircraft transactions. *Journal of Finance*, 53(3):939–978.

Roukny, T., Bersini, H., Pirotte, H., Caldarelli, G., and Battiston, S. (2013). Default cascades in complex networks: topology and systemic risk. *Scientific Reports*, 3(1):2759.

Saracco, F., Di Clemente, R., Gabrielli, A., and Squartini, T. (2015). Randomizing bipartite networks: the case of the world trade web. *Scientific Reports*, 5(10595).

Schaanning, E. F. (2016). Fire sales and systemic risk in financial networks. *PhD thesis, Imperial College London*.

Shleifer, A. and Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, 25(1):29–48.

Siemsen, T. and Vilsmeier, J. (2018). On a quest for robustness: about model risk, randomness and discretion in credit risk stress tests. *Deutsche Bundesbank Discussion Paper*, 31.

Squartini, T., Almog, A., Caldarelli, G., Van Lelyveld, I., Garlaschelli, D., and Cimini, G. (2017). Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks. *Physical Review E*, 96.

Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A., and Garlaschelli, D. (2018). Reconstruction methods for networks: the case of economic and financial systems. *Physics Reports*, 757:1–47.

Squartini, T. and Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):83001.

Swets, J. A., Dawes, R. M., and Monahan, J. (2000). Better decisions through science. *Scientific American*, 283(4):82–87.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125.

Vitali, S., Glattfelder, J. B., and Battiston, S. (2011). The network of global corporate control. *PLoS ONE*, 6(10):e25995.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1):32–35.

Zhang, P., Wang, J., Li, X., Li, M., Di, Z., and Fan, Y. (2008). Clustering coefficient and community structure of bipartite networks. *Physica A*, 387(27):6869–6875.