

Discussion Paper

Deutsche Bundesbank
No 45/2020

Backtesting macroprudential stress tests

Amanah Ramadiah

(University College London)

Daniel Fricke

(University College London and Deutsche Bundesbank)

Fabio Caccioli

(University College London, London School of Economics and Political Science
and London Mathematical Laboratory)

Editorial Board:

Daniel Foos
Stephan Jank
Thomas Kick
Malte Knüppel
Vivien Lewis
Christoph Memmel
Panagiota Tzamourani

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet <http://www.bundesbank.de>

Reproduction permitted only if source is stated.

ISBN 978-3-95729-753-2 (Printversion)

ISBN 978-3-95729-754-9 (Internetversion)

Non-technical summary

Research Question

Different macroprudential stress tests were proposed in the literature, which account for potential contagion effects in financial networks. While these models were useful for building intuition about how shocks may propagate through the system, their capability to accurately predict whether a given bank will default has not been the focus of the literature. It is well-known that different models may yield very different stress testing results. We therefore propose a backtesting framework that assesses the predictive performance of different fire-sale stress test models and allows to choose the most accurate model from a set of alternatives.

Contribution

We introduce a generalized fire-sale stress test model that captures a wide range of behavioral assumptions with regards to banks' liquidation dynamics under stress. The literature has proposed alternative behavioral assumptions in this regard, all of which are covered by our generalized model. We build a network of common asset holdings using public balance-sheet data for U.S. commercial banks in 2007. We then compare the model predictions with the list of actual defaults that occurred in the U.S. during the years 2008-2010. In order to assess the relative performance of these network models, we also use several alternative benchmarks.

Results

We identify two asset classes for which the model has predictive power, independently of the assumed liquidation dynamics. We then show how the behavioural assumption yielding the most accurate model depends on the size of the initial shock and on secondary market liquidity. We also identify, for different liquidation dynamics, the optimal number of liquidation rounds. Overall, our analysis shows that properly calibrated macroprudential stress tests can have predictive power superior to alternative benchmarks that do not account for the network of common asset holdings.

Nichttechnische Zusammenfassung

Fragestellung

In der Literatur wurden verschiedene makroprudenzielle Stresstests vorgeschlagen, welche mögliche Ansteckungseffekte aufgrund von Netzwerkverbindungen zwischen Finanzintermediären berücksichtigen. Diese Ansätze haben unser Verständnis verbessert, ob und wie sich Schocks in solchen Netzwerken ausbreiten können, doch wurde bisher kaum untersucht, inwiefern diese Ansätze den Ausfall von Kreditinstituten vorhersagen können. Aus der Literatur ist bekannt, dass verschiedene Stresstest-Methoden sehr unterschiedliche Ergebnisse liefern können. Wir schlagen einen Backtesting-Ansatz vor, der die Güte der Vorhersagen verschiedener Ansätze quantifiziert und aus verschiedenen Ansätzen das präziseste Modell auswählt.

Beitrag

Wir schlagen ein verallgemeinertes Stresstest-Modell für Notverkäufe vor, welches verschiedene Verhaltensannahmen bezüglich des Liquidationsverhaltens von Banken im Stressfall berücksichtigen kann. In der Literatur wurden diesbezüglich unterschiedliche Verhaltensannahmen untersucht, welche alle von unserem Modell abgedeckt werden. Basierend auf öffentlichen Bilanzdaten konstruieren wir das Netzwerk überlappender Wertpapier-Portfolios zwischen US-Banken im Jahr 2007. Wir vergleichen die Prognosen des Modells mit den tatsächlichen Ausfällen von US-Banken in den Jahren 2008 bis 2010. Um die relative Prognosegüte der Modelle zu bestimmen, nutzen wir auch alternative Prognoseansätze, welche das Netzwerk überlappender Bankenportfolios nicht berücksichtigen.

Ergebnisse

Wir identifizieren zwei Klassen von Wertpapieren für welche das Netzwerkmodell besondere Prognosegüte aufweist, unabhängig von der jeweils getroffenen Verhaltensannahme bzgl. des Liquidationsverhaltens von Banken. Wir zeigen, dass die Prognosegüte der Modelle von der Stärke des Ausgangsschocks und von der Sekundärmarktliquidität abhängt. Zudem hängt die optimale Anzahl an Liquidationsrunden des Modells wiederum vom unterstellten Liquidationsverhalten ab. Insgesamt zeigen unsere Analysen, dass angemessen kalibrierte makroprudenzielle Stresstests eine bessere Prognosegüte erzielen können als Ansätze ohne Netzwerkkomponente.

Backtesting Macroprudential Stress Tests*

Amanah Ramadiah^{1†}, Daniel Fricke^{1,2}, and Fabio Caccioli^{1,3,4}

¹ University College London, Department of Computer Science, United Kingdom

² Deutsche Bundesbank, Directorate General Financial Stability, Germany

³ London School of Economics and Political Science, Systemic Risk Centre, United Kingdom

⁴ London Mathematical Laboratory, United Kingdom

August 3, 2020

Abstract

In this paper, we consider models of price-mediated contagion in a banking network of common asset holdings. For these models, the literature proposed two alternative classes of liquidation dynamics: *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). We introduce a one-parameter family of non-linear liquidation functions that interpolates between these two extremes. We then test the capability of these models to predict actual bank defaults (and survivals) in the United States for the years 2008–10. We show that the model performance depends on the type of shock being imposed (idiosyncratic versus systematic). We 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. We 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.

Keywords: systemic risk; fire sales; price-mediated contagion; common asset holdings.

JEL Classification: G01, G11

*We thank participants of 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. A.R. acknowledges a PhD scholarship from the Indonesia Endowment Fund for Education (LPDP). F.C. acknowledges the support of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (ES/K002309/1 and ES/R009724/1). The views expressed in this paper represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

[†]Corresponding author. Email: a.ramadiah@cs.ucl.ac.uk

1 Introduction

The interconnectedness of the financial system has been considered by many as an important source of systemic risk (Haldane and May (2011)). Ever since the global financial crisis of 2008-09, the literature on network models of the financial system has grown rapidly (see Glasserman and Young (2016) for a recent survey). Much of the literature has focused on the development of contagion network models and various stress test models based on different assumptions have been introduced (Baranova et al. (2017); Bardoscia et al. (2016); Battiston et al. (2012); Caccioli et al. (2014); Cont and Schaanning (2017); Duarte and Eisenbach (2015); Eisenberg and Noe (2001); Elsinger et al. (2006); Fricke and Fricke (2020); Greenwood et al. (2015); Huang et al. (2013)). 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 survivals) during crisis periods. 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 dynamics that were observed during the most recent crisis period. This is particularly important given that a multitude of models have 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 outcomes of stress tests crucially hinge upon the underlying model (e.g., Niepmann and Stebunovs (2018); Siemsen and Vilsmeier (2018)). Our paper contributes to the literature by providing a structured approach to picking the desired model -based on the researcher's/policymakers' specific criteria of interest- from a set of available options.

Broadly speaking, the existing literature considers mainly two contagion channels for banking/financial networks: (i) direct contagion in the interbank borrowing/lending network, and (ii) indirect price-mediated contagion due to common asset holdings (overlapping portfolios). Here we look at the second contagion channel, which appears to be more relevant in general (Caccioli et al. (2014); Glasserman and Young (2015)). This contagion mechanism 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. Two extreme types of banks' liquidation dynamics have been proposed in the literature, namely the model of Caccioli et al. (2014) and Huang et al. (2013) and 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 acknowledge that the actual liquidation behaviour of banks might lie somewhere in-between these extremes, we propose a fire sales 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.

We then use the model to predict actual bank defaults for a range of liquidation dynamics, i.e., different values of γ . Our goal is to identify the value(s) of γ that perform(s) best in terms of predictive accuracy with regards to actual bank defaults and survivals. Following [Huang et al. \(2013\)](#), we use U.S. commercial bank balance sheet data for the last quarter of 2007, and we apply a shock that is meant to mimic the onset of the subprime crisis. We 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).

Our main findings are as follows: the performance of the stress test 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. Our approach allows to identify those asset classes that appear most relevant. Based on this identification, we find that the stress test model displays a 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 that does not account for the network of common asset holdings: with leverage and total assets as the sole explanatory variables. Furthermore, we find that the best performing liquidation dynamics depend on the other model parameters, namely the size of the initial shock and the level of market liquidity. We 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 macroprudential stress tests. Lastly, we 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, we show that considering only the first round of asset liquidations appears most accurate for a model with small γ (banks act as leverage targeters), while accounting for multiple rounds of asset liquidations provides better results for larger γ (banks only liquidate in case of default).

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

we wish to match are individual defaults/non-defaults.

The remainder of this paper is structured as follows: in section 2 we present the model. Section 3 provides details on the data sets being used in the paper. Section 4 contains the model application and section 5 concludes.

2 Model dynamics

We consider a financial network of N banks and M assets. A link exists between bank i and asset j if i invests in j . We should note that the network is bipartite, which implies the absence of inter-bank and inter-asset links.

Each bank i holds a portfolio of risky assets $\{Q_{i,1}^t, \dots, Q_{i,M-1}^t\} \geq 0$ and cash $Q_{i,M}^t = C_i^t \geq 0$ at time t . Let f_j^t be the price of asset j at time t , with the price of cash remaining constant over time ($f_M^t = f_M^0 \forall t$), which we assume to be discrete. The total value of bank i 's investment portfolio is therefore:

$$A_i^t = \sum_{j=1}^M Q_{ij}^t f_j^t. \quad (1)$$

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

$$A_i^t \equiv E_i^t + L_i^t, \quad (2)$$

and bank i does not default at time t if:

$$A_i^t \geq L_i^t. \quad (3)$$

The *weight* of asset j in bank i 's portfolio is defined as:

$$w_{ij}^t = \frac{Q_{ij}^t f_j^t}{A_i^t}. \quad (4)$$

The *total number of outstanding shares* of asset j held in the system is denoted as:

$$\beta_j = \sum_{i=1}^N Q_{ij}^0. \quad (5)$$

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

$$\lambda_i^t = \frac{A_i^t}{E_i^t}, \quad (6)$$

and λ_i^0 is bank's i initial leverage. λ_i^0 will be of particular interest for the *leverage targeting model*.

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

1. We impose an initial shock on the value of some asset(s).
2. Banks update the value of their portfolios. A decline in the values of its assets can lead banks to liquidate part of their holdings, if they target their leverage ratios to some extent. If the decline is large enough to make the bank default, it will liquidate all its holdings. In the threshold model, asset liquidations only occur in the case of default.
3. Asset liquidations generate price impact, so the value of an asset is recomputed depending on the volume of the asset that has been liquidated.
4. Back to step 2 (optional).

Let us discuss these steps in more detail.

2.1 Step 1: Initial shock

Suppose we initially impose a shock on asset j by reducing its value to a fraction $p \in [0, 1]$ of its original value:

$$f_j^1 = p_j^0 f_j^0. \quad (7)$$

Note that 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:

$$A_i^1 = A_i^0 - Q_{ij}^0 (1 - p_j^0) f_j^0. \quad (8)$$

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

$$\phi_i^1 = -\frac{(A_i^1 - A_i^0)}{A_i^0} \quad (9)$$

and bank i 's updated equity becomes:

$$E_i^1 = E_i^0 - \phi_i^1 A_i^0, \quad (10)$$

while its liabilities remain unchanged, $L_i^1 = L_i^0$.

Note that bank i 's updated leverage ratio reads as:

$$\lambda_i^1 = \frac{(1 - \phi_i^1) A_i^0}{(1 - \phi_i^1) E_i^0 - \phi_i^1 L_i^0} \geq \lambda_i^0. \quad (11)$$

Hence, in response to a drop in asset prices, leverage will mechanically increase when

liabilities remain fixed (Adrian and Shin (2010)). Finally, bank i has not defaulted if

$$\phi_i^1 \leq \frac{1}{\lambda_i^0}. \quad (12)$$

2.2 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 (12) 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. (11), 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:

$$\Pi_i^1 = A_i^1 \left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1} \right). \quad (13)$$

To accommodate for the case where bank i defaults and needs to sell all of its remaining assets, we modify the function Π_i^1 in Equation (13) as follows:

$$\Pi_i^1 = A_i^1 \min \left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1}, 1 \right). \quad (14)$$

In contrast to the leverage targeting model, the *threshold model* (Caccioli et al. (2014); Huang et al. (2013)) assumes that bank i will only liquidate assets when it defaults. One might hypothesize that actual bank behaviour might lie in-between these two extreme behaviours. Therefore we introduce a one-parameter non-linear functional form, $G_i^1(\phi)$ that interpolates between them. In order to derive the functional form of $G_i^1(\phi)$, we start by looking at Equation (14), which describes the behaviour of bank i in the leverage targeting model. We want to modify the equation such that it also encompasses the behaviour of bank i in the threshold model. We do this by incorporating $G_i^1(\phi)$ in Equation (14), such that:

$$\Pi_i^1 = G_i^1(\phi) A_i^1 \min \left(1 - \frac{\lambda_i^0 E_i^1}{A_i^1}, 1 \right). \quad (15)$$

In the leverage targeting model we have:

$$G_i^1(\phi) = 1, \quad (16)$$

while in the threshold model:

$$G_i^1(\phi) = \begin{cases} 0, & \text{if } \phi_i^1 \leq \frac{1}{\lambda_i^0}, \\ 1, & \text{otherwise.} \end{cases} \quad (17)$$

To interpolate between leverage targeting and threshold dynamics, we introduce the function $G_i^1(\phi)$:

$$G_i^1(\phi) = \min \left[e^{\gamma \left(\phi_i^1 - \frac{1}{\lambda_i^0} \right)}, 1 \right], \quad (18)$$

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 $\gamma = 0$, we 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 $\gamma \rightarrow \infty$, we recover the threshold model. This is shown in [Figure 1](#), where we present a comparison between different values of γ in term of liquidation volume as a function of absolute return.

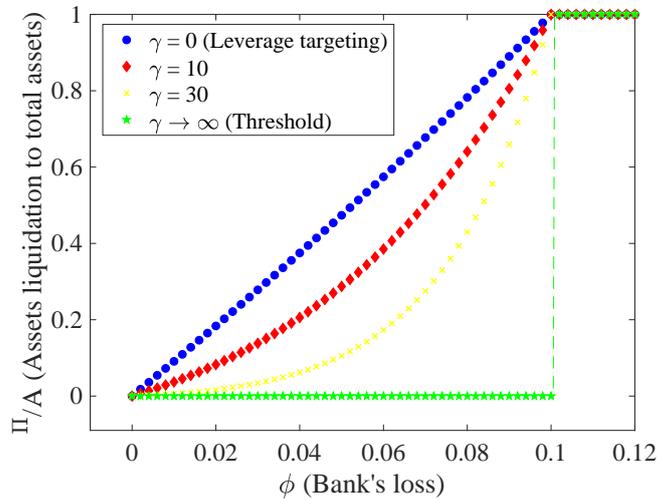


Fig. 1: The volume of assets liquidated (relative to total assets) as a function of the loss in value of assets for a portfolio with an initial leverage of 10, for different values of γ in the model.

Following the liquidation procedure, we recompute the total assets of bank i as:

$$A_i^{1+\frac{1}{2}} = A_i^1 - \Pi_i^1. \quad (19)$$

Notice that we use time $t = 1 + \frac{1}{2}$ for the new total assets in [Eq. \(11\)](#). This is to clarify that the change is due to an intermediate step between time $t = 1$ (when the bank liquidates part of its asset) and time $t = 2$ (when asset prices decrease because of the liquidation), which we will discuss next.

2.3 Step 3: Fire sales generate price impact

We previously introduced Π_i^1 , the total volume of assets liquidated by bank i . Now we specify, π_{ij}^1 , the asset-specific liquidation volumes of bank i . In line with the literature ([Cont and Schaanning \(2017\)](#); [Fricke and Fricke \(2020\)](#); [Greenwood et al. \(2015\)](#)),

banks are assumed to liquidate assets proportional to their current portfolio weights (pro rata liquidation). This assumption is supported by empirical studies on asset sales of banks and investment funds during market stress (e.g., Jiang et al. (2017)).¹ Therefore, bank i sells a total number of shares asset j equal to:

$$\pi_{ij}^1 = \frac{w_{ij}^1 \Pi_i^1}{f_j^1}. \quad (20)$$

In line with empirical evidence, this additional selling pressure generates market impact, which leads to a further devaluation of the assets. Let x_j^1 be the fraction of asset j that has been liquidated at time $t = 1$,

$$x_j^1 = \frac{\sum_{i=1}^N \pi_{ij}^1}{\beta_j}. \quad (21)$$

The market price of asset j at time $t = 2$ is updated to:

$$f_j^2 = \Psi_j(x_j^1) f_j^1, \quad (22)$$

where Ψ_j is the market impact function of asset j . Given that prices react to asset liquidations, this will cause further mark-to-market losses to banks. We should note that banks can also “liquidate” cash (banks use cash to repay existing debt), but without any market impact.

The assumed functional form of Ψ_j varies across different stress test models (see Appendix A.1 for classification of existing fire sales models), and we refer the interested reader to Cont and Schaanning (2017) for a comprehensive discussion of this topic. Similarly, many studies make a simplifying homogeneity assumption with regards to the market impact function (e.g. Greenwood et al. (2015)). Put simply, this means that all assets are assumed to be equally liquid, i.e., $\Psi_j = \Psi \forall j$. In line with the literature, we therefore use a linear market impact function and make the homogeneity assumption in our main specification. Formally, 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, that is: $\Psi_j(x) = 1 - \alpha x_j^1$. The price of asset j at time $t = 2$ then becomes:

$$f_j^2 = (1 - \alpha x_j^1) f_j^1, \quad (23)$$

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 would 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.

Later on, we will relax the homogeneity assumption (see section 4.4) and the linearity assumption (see Appendix).

¹An alternative approach would consist of banks selling the most liquid assets first (waterfall liquidation). In a setting with homogeneous price impacts, however, the two approaches are equivalent.

2.4 Step 4: Back to step 2 (optional)

We have described the first complete iteration round of the model, which results in asset liquidations generating price impact. Following this 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 Data

As usual, once a stress test model has been developed, it is calibrated on empirical data sets to perform counter-factual simulations regarding the stability of the underlying financial networks (Baranova et al. (2017); Bardoscia et al. (2016); Battiston et al. (2012); Cont and Schaanning (2017); Di Gangi et al. (2018); Duarte and Eisenbach (2015); Elsinger et al. (2006); Fricke and Fricke (2020); Greenwood et al. (2015); Huang et al. (2013); Levy-Carciente et al. (2015); Ramadiah et al. (2020); see also Upper (2011) for related works).²

Following Huang et al. (2013), our data come from two sources: first, we take U.S. commercial banks' balance sheet data from Wharton Research Data Services³ 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 1 for a list of asset classes). This gives us the empirical equivalent of matrix Q , which is the bipartite financial network with dimension $7,783$ banks \times 14 assets. We also have information on the total assets, total liabilities, and equity of each bank. Note that most of the asset classes considered here are loans, and therefore are clearly illiquid. However, we should stress that many of these can indeed be traded on secondary markets (see, e.g., Drucker and Puri (2009)).⁴

The second data set is a list of bank failures from the Federal Deposit Insurance Corporation (FBL-FDIC)⁵ for the period 1/1/2008 - 7/1/2011. During this period a total number of 370 banks failed; for 306 of these we have the corresponding balance sheet data from the first database. Hence, roughly 4% of the banks in our sample defaulted during the financial crisis period.

In the following, we apply our stress test model to the bipartite network from the first data set. We then compare the list of banks that the model predicts to default and the actual list of bank failures. We 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.

Fundamentally, our paper is related to the growing literature on reverse stress test-

²Alternatively, theoretical analyses are performed on stylized synthetic data sets to study the effect of particular financial networks topologies (e.g., Nier et al. (2007)).

³<https://wrds-web.wharton.upenn.edu/wrds/>

⁴For example, Keys et al. (2010) note that the market of securitized mortgage loans reached \$3.6 trillion prior to the crisis.

⁵<https://www.fdic.gov/bank/individual/failed/banklist.html>

Index	Asset class	Total amount (million \$)	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	Obligations (other than securities and leases) of states and political subdivision in the U.S.	282.80	2.23
10	Loans secured by nonfarm nonresidential properties	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 1: Overview of the asset classes used in the exercise. 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 million \$), and the fourth column is the corresponding percentage share of total assets.

ing (e.g., [Grigat and Caccioli \(2017\)](#)), which aims to identify (stress) scenarios that are compatible with a certain stress testing outcome. Here the outcome is the default/non-default of individual banks. From this perspective, while one might be worried that our approach may be prone to overfitting, we argue that our approach is not actually concerned with fitting the optimal parameter γ in an statistical/econometric sense. Rather, we wish to find the value(s) of γ that replicate(s) as closely as possible the behavior of U.S. banks during the crisis. We will return to this point in [Section 4.3](#). Of course, future research should explore to what extent the findings vary for different countries and for different time periods.

4 Results

4.1 Experimental setup

Let us spell out some assumptions and introduce some definitions that we use in this study. All assets start out with the same initial price:

$$f_j^0 = 1, \forall j \in \{1, \dots, M\},$$

and we start the cascading process by reducing the value of an asset $j \in \{1, \dots, M-1\}$ to a fraction p of its original value ($p \in \{0, 0.01, 0.02, \dots, 0.99, 1\}$). In this sense, a large (small) value of p corresponds to small (large) initial shock. With regards to the market impact function, Ψ , we iterate over different values of α ($\alpha \in \{0, 0.01, 0.02, \dots, 0.99, 1\}$) and, initially, we assume that this value is homogeneous across assets (except for cash,

for which we always consider $\alpha = 0$). We will relax the homogeneity assumption below.

In the following, we will show results for all possible combinations of p and α , but we mainly focus on the results for a restricted range of parameters. Since the majority of asset classes considered here are relatively illiquid, we follow [Ramadiah et al. \(2020\)](#) and focus on the upper range of the market impact parameter ($\alpha \in [0.6, 1]$).⁶ Moreover, in line with previous studies on price-mediated contagion ([Ramadiah et al. \(2020\)](#), [Cont and Schaanning \(2017\)](#), [Greenwood et al. \(2015\)](#)), we 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, we 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 2 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. We later compare the model accuracy when allowing for multiple rounds of asset liquidations.

Concerning the propensity of banks to follow threshold dynamics (parameter γ), we consider values of $\gamma \in \{0, 1, \dots, 49\} \cup \{\infty\}$. Recall that a value of $\gamma \rightarrow \infty$ 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 our set of values, we perform a stress test and measure the prediction accuracy of the corresponding model. To this end, we evaluate each model's ability to identify bank failures and non-failures during financial distress correctly. Here we 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 diagonal line (from bottom left to top right) in the ROC space; see [Figure 2](#). In contrast, a perfect model produces points closer to the top left corner in the ROC space.

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 we use *Youden's J statistic* ([Youden \(1950\)](#)), which is defined as :

$$J_{\gamma,p,\alpha} = \text{TPR}_{\gamma,p,\alpha} - \text{FPR}_{\gamma,p,\alpha}. \quad (24)$$

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 2](#) illustrates the computation of J .

Unless otherwise stated, in our analysis below we 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

⁶For $\alpha = 0.6$, 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.

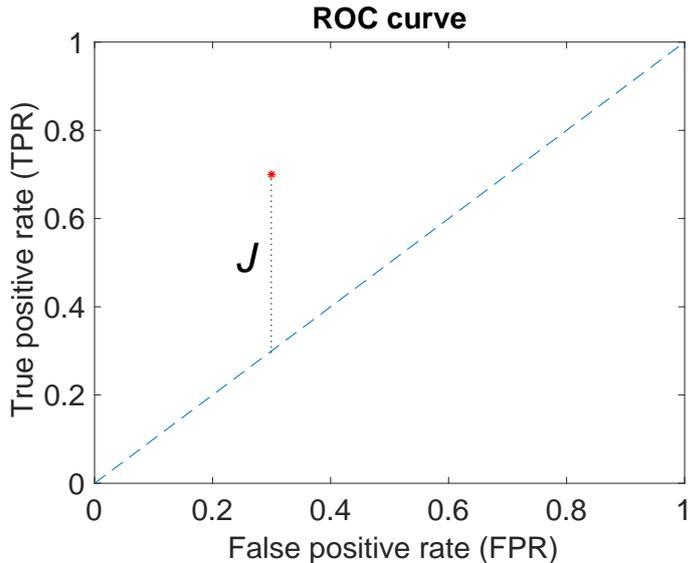


Fig. 2: Illustration of Youden’s J statistic. The red point in the ROC space corresponds to $\text{FPR} = 0.3$ and $\text{TPR} = 0.7$. The value of J of this point is displayed by the dotted line ($J = 0.7 - 0.3 = 0.4$). Larger values indicate a better model performance.

from the influence of the initial shock. (In the Appendix, we take a closer look at defaults due to the initial shock.) Moreover, to compare the accuracy of the network models with some benchmarks, we 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 We 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 4.3 we will introduce a more sophisticated benchmark model that also includes network information.)

We 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, we consider failures of all banks in the data set, while in the latter we only consider failures due to the contagion process and exclude those due to the initial shock. Our 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. We should highlight that if we also considered failures due to the initial shock for the network model, we would obtain more superior ROC curves compared to those that we show here.⁸ Therefore, if anything, in comparison with the logistic regressions, the accuracy of the network model is biased downwards.

4.2 Model performance

Apart from the model-specific parameters (in our 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)

⁷See Appendix A.2 for details on the regression model.

⁸See Appendix A.4

should be shocked. In the following, we will show how to pick the *relevant* asset classes in terms of how the stress test model performs in the classification exercise. We 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 we 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 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)).

4.2.1 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 3, this yields very poor classification results. Specifically, we run the stress test 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 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 class (j)	Complete parameter range					Restricted parameter range				
	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 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, we show separate results for the individual asset classes in Table 2. In particular, we compute the average value of *Youden's J statistic* (Equation (24)) 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 statistics, we also perform a simple simulation-based significance test: first, we 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 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 , we can then count how many of the observations fall within the 5% confidence interval of the random classifier.

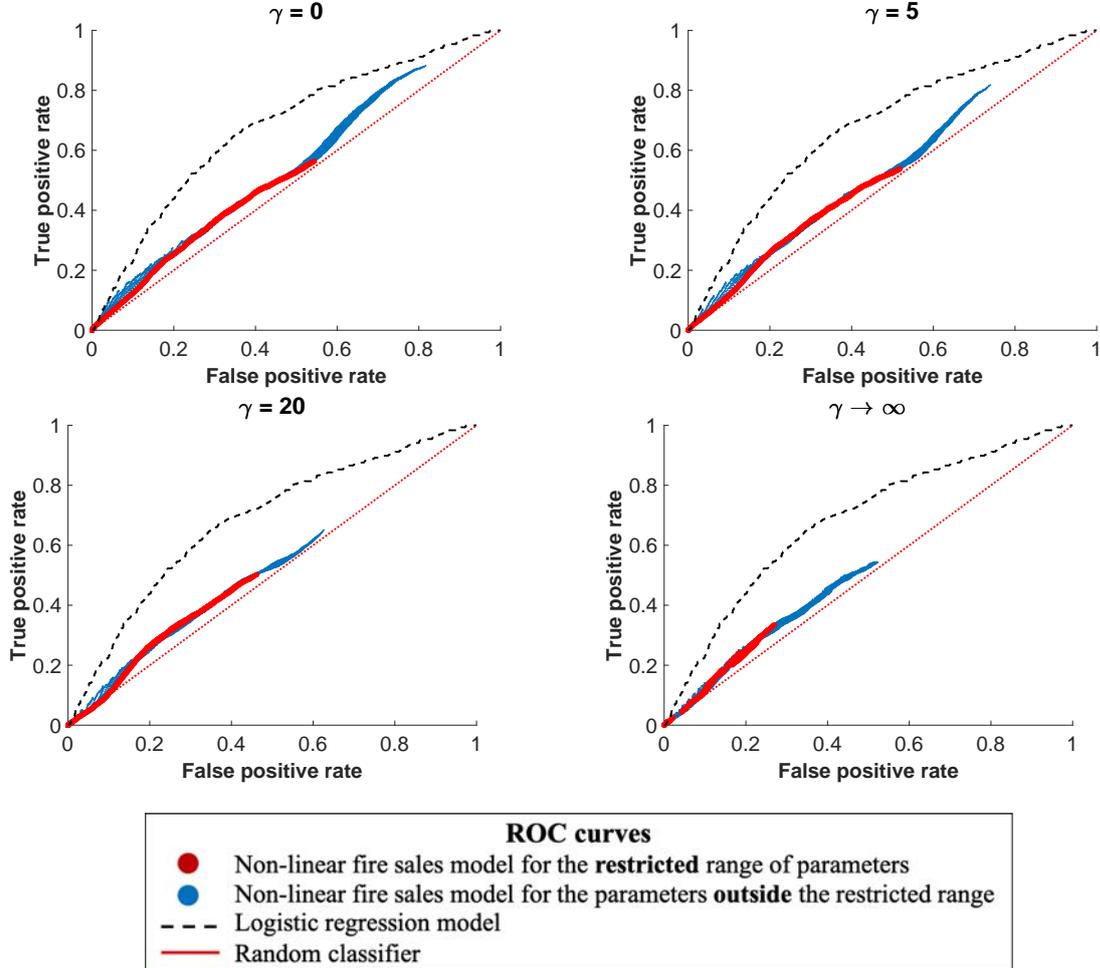


Fig. 3: **ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., for the average results of initial shock across all asset classes (except cash).** Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

This number is reported as Prob in Table 2. Based on this simple significance test, we 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, we 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.⁹ Table 2 also shows that for asset classes 1 and 5 the unconditional results are slightly better: approximately 60% of the J statistics fall outside the 5% confidence bands of the random classifier. While these findings would still not count as statistically significant at reasonable confidence levels, we treat these two asset classes as the most relevant ones in what follows.

Note that our 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, we find that the classifications are significantly better compared to those of a random classifier for initial shocks on these two asset classes. We 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).

4.2.2 Results for the two most relevant asset classes

The different panels in Figure 4 show the ROC curves based on $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, when imposing an initial shock on asset class 1. Figure 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, blue dots indicate results outside this range. Overall, we 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 Figures Figure 4 and Figure 5, where the ROC curves lie above diagonal lines. In most instances, we also find that the model is 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 4 and Figure 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

⁹We 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 we reject the null hypothesis for all asset classes.

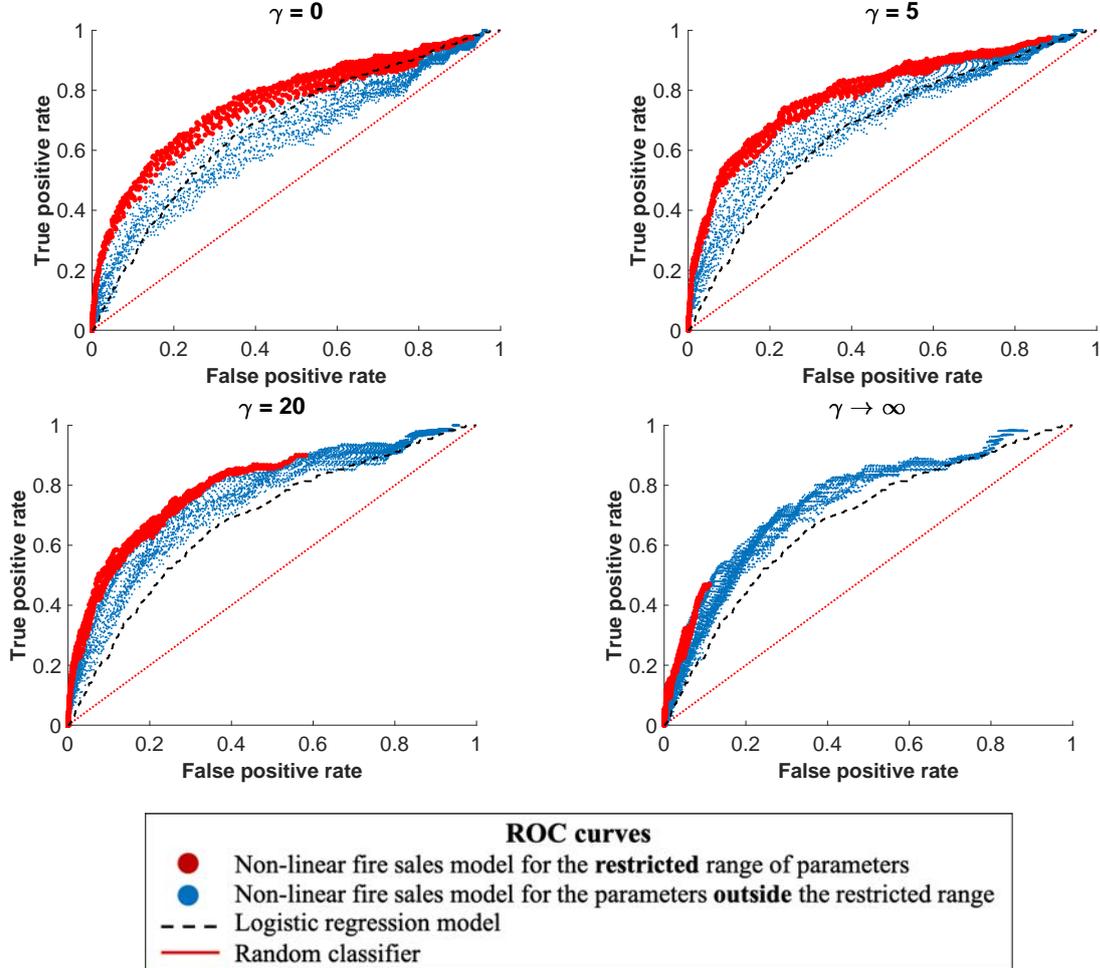


Fig. 4: **ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with initial shock on loans for construction and land development (asset class 1).** Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

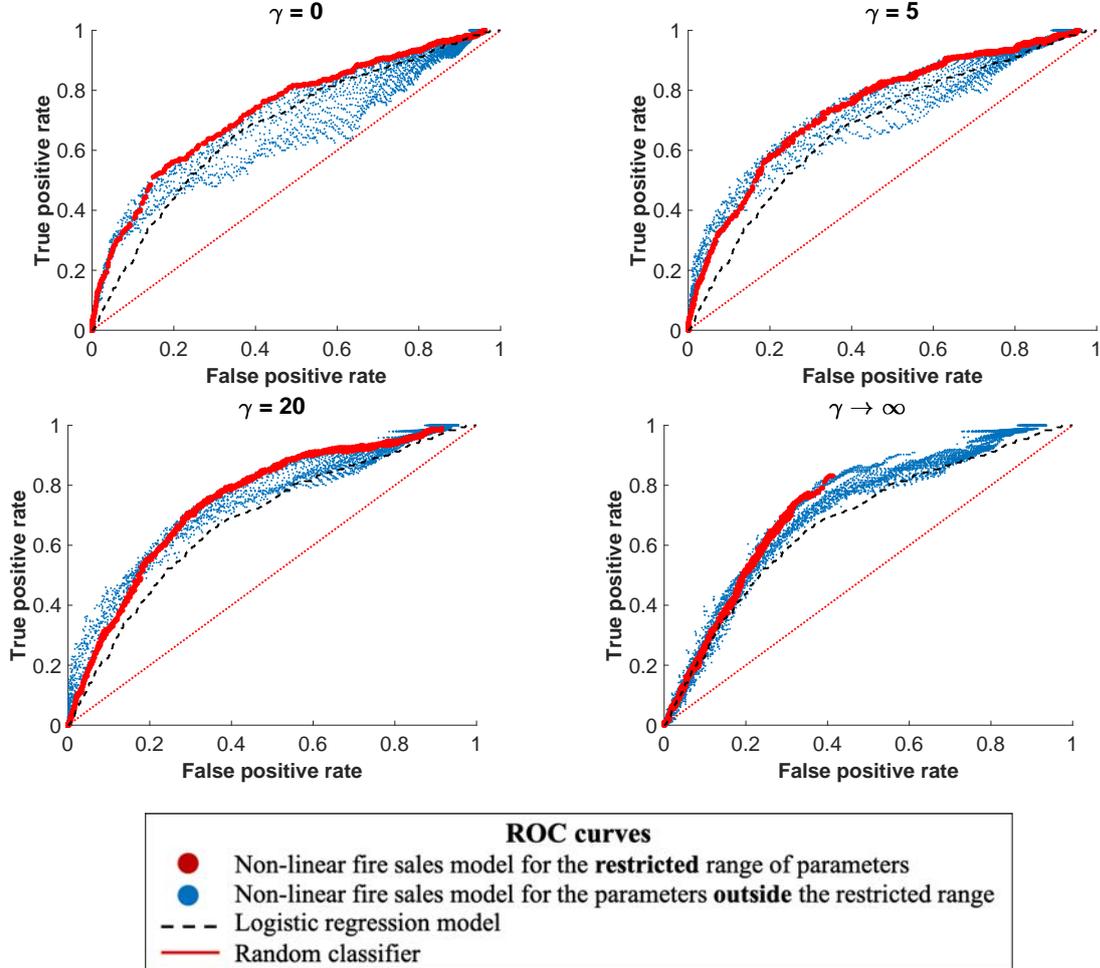


Fig. 5: **ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with initial shock on loans secured by non-farm non-residential properties (asset class 5).** Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

(see section 4.4). As we see from the top panels in Figure 4 and Figure 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 performing.

4.2.3 Model performance in the (p, α) plane

In the following, we 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, we find the value of γ for which $J_{\gamma, p, \alpha}$ is maximized:

$$\gamma^{\text{opt}}(p, \alpha) = \arg \max_{\gamma} J_{\gamma, p, \alpha}. \quad (25)$$

As noted previously, here we focus only on the results for the two most relevant asset classes (1 and 5). In the following, we 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 6(a) and Figure 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 ($\gamma^{\text{opt}}(p, \alpha) \rightarrow \infty$).¹⁰ 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 ($\gamma^{\text{opt}}(p, \alpha) \rightarrow 0$). 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 6(b) and Figure 7(b) show the value of $J_{\gamma^{\text{opt}}, p, \alpha}$ in the (p, α) plane. As we 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 6(c) and Figure 7(c) show the differences in performance between the best and the worst model in the (p, α) plane:

$$\delta = \max_{\gamma} J_{\gamma, p, \alpha} - \min_{\gamma} J_{\gamma, p, \alpha}, \quad (26)$$

¹⁰We also see a few cases in region 1 where the accuracy of all values of γ is equivalent (shown in white).

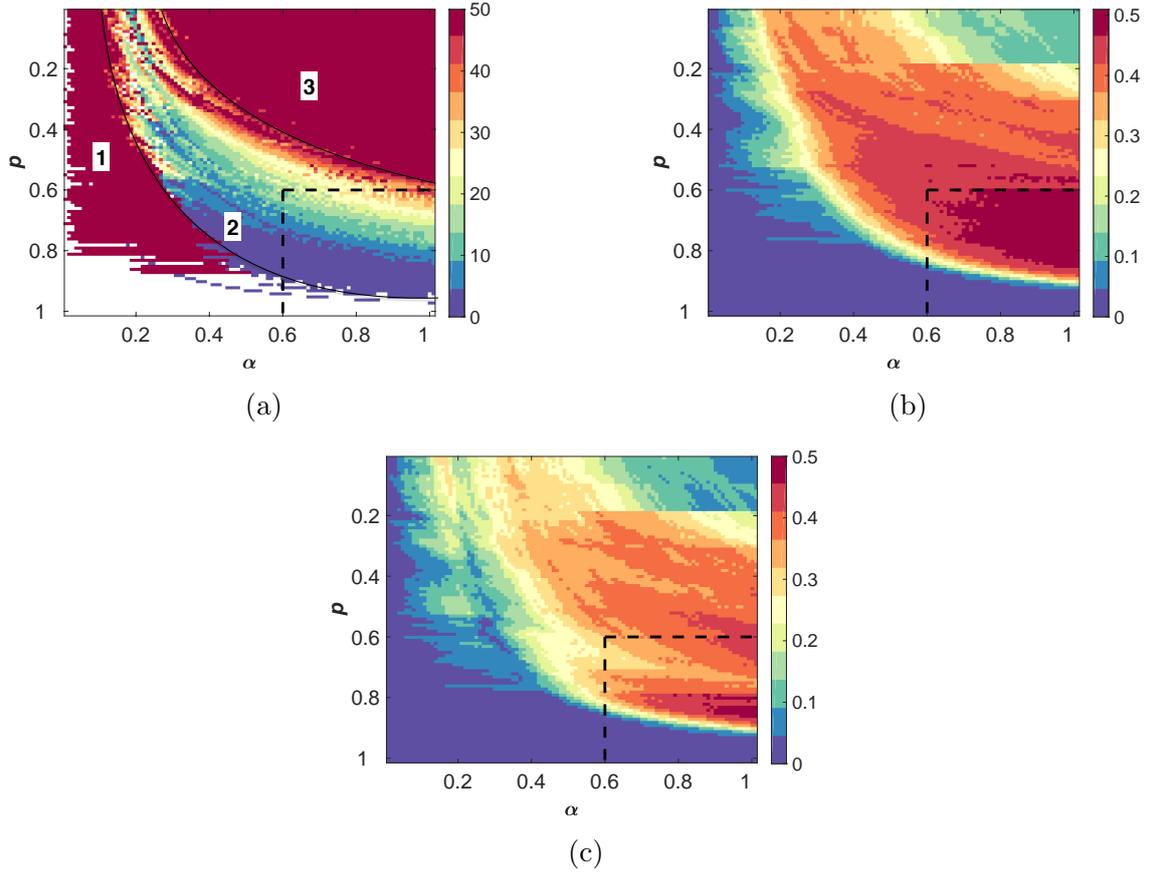


Fig. 6: **Results from imposing an initial shock on asset class 1 (loans for construction and land development).** (a) The variation of γ^{opt} (based on Youden's J statistic) in different (p, α) regimes. Warmer (colder) colors refer to larger (smaller) values of γ , and white is used whenever the accuracy of all γ are equivalently similar. We look at the range of $\gamma \in \{0, \dots, 49\} \cup \{\infty\}$, and $\gamma = 50$ in this plot corresponds to $\gamma \rightarrow \infty$. (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.

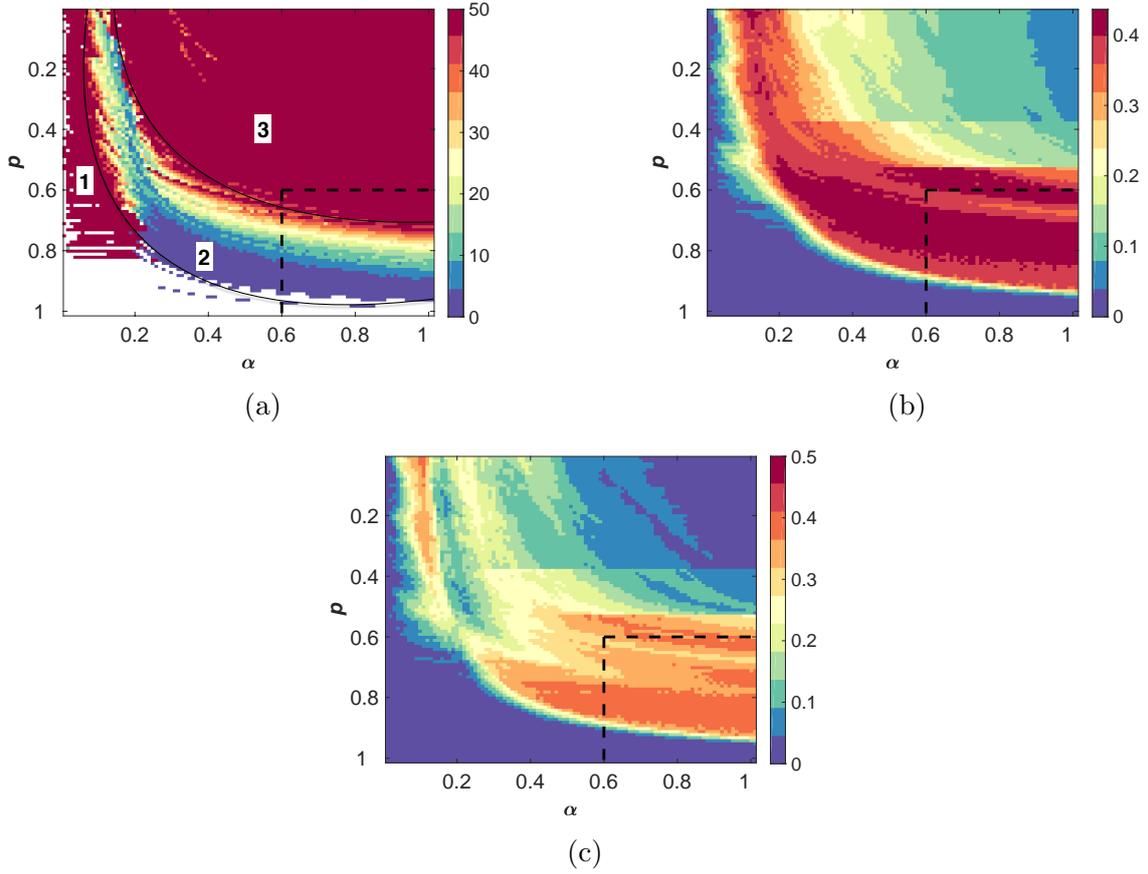


Fig. 7: **Results from imposing an initial shock on asset class 5 (loans secured by non-farm non-residential properties).** (a) The variation of γ^{opt} (based on Youden's J statistic) in different (p, α) regimes. Warmer (colder) colors refer to larger (smaller) values of γ , and white is used whenever the accuracy of all γ are equivalently similar. We look at the range of $\gamma \in \{0, \dots, 49\} \cup \{\infty\}$, and $\gamma = 50$ in this plot corresponds to $\gamma \rightarrow \infty$. (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, we 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 our main findings can be found in Table 3. In 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 we provided in Figure 6 and Figure 7.

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

Table 3: The best performing liquidation dynamics for different combinations of the exogenous shock and the market impact parameter.

Restricted Parameter Range. We 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 6 and Figure 7 as the area inside the black dashed line square (bottom right corner). Interestingly, Figure 6(a) and Figure 7(a) show that the best performing liquidation parameter γ in this range lies in-between leverage targeting and the threshold dynamics. Furthermore, we also observe from Table 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 ($\gamma \rightarrow \infty$). 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 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.

4.3 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. An obvious question is how well the model performs relative to more sophisticated benchmarks. For this purpose, we specify an enhanced logistic regression, which also includes information on the network of common asset holdings. Specifically, we include each bank’s relative amount of holdings of each asset class (defined as the portfolio share).¹¹ As shown in the regression results in Table A.3 in the Appendix, the parameter on $\log(\text{Leverage})$ now turns out insignificant. Only the parameters on $\log(\text{TotalAssets})$, and on asset classes 1 and 8 are significant. Nevertheless, the pseudo- R^2 now increases substantially from 0.01 to 0.17.

In Figures 8 and 9, we plot the performance of this benchmark against that of the network model when we 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 flexibility of the network model can be seen in Figure 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

¹¹To deal with collinearity, we exclude asset classes 10 and 13 in the regression, which account only for less than 2% of the aggregated balance sheet.

the same banks to default in both cases.

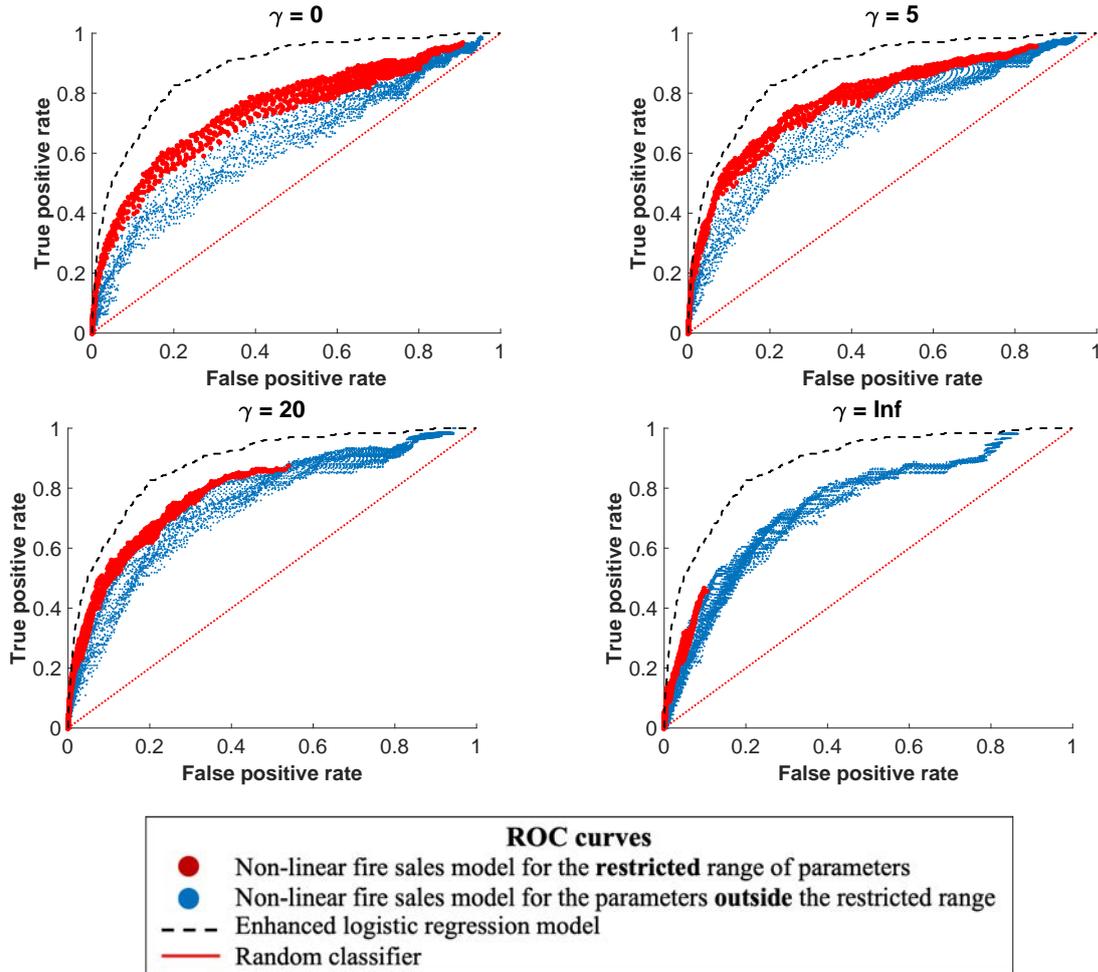


Fig. 8: **ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with initial shock on loans for construction and land development (asset class 1).** Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

4.4 Extensions

In the following, we 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.

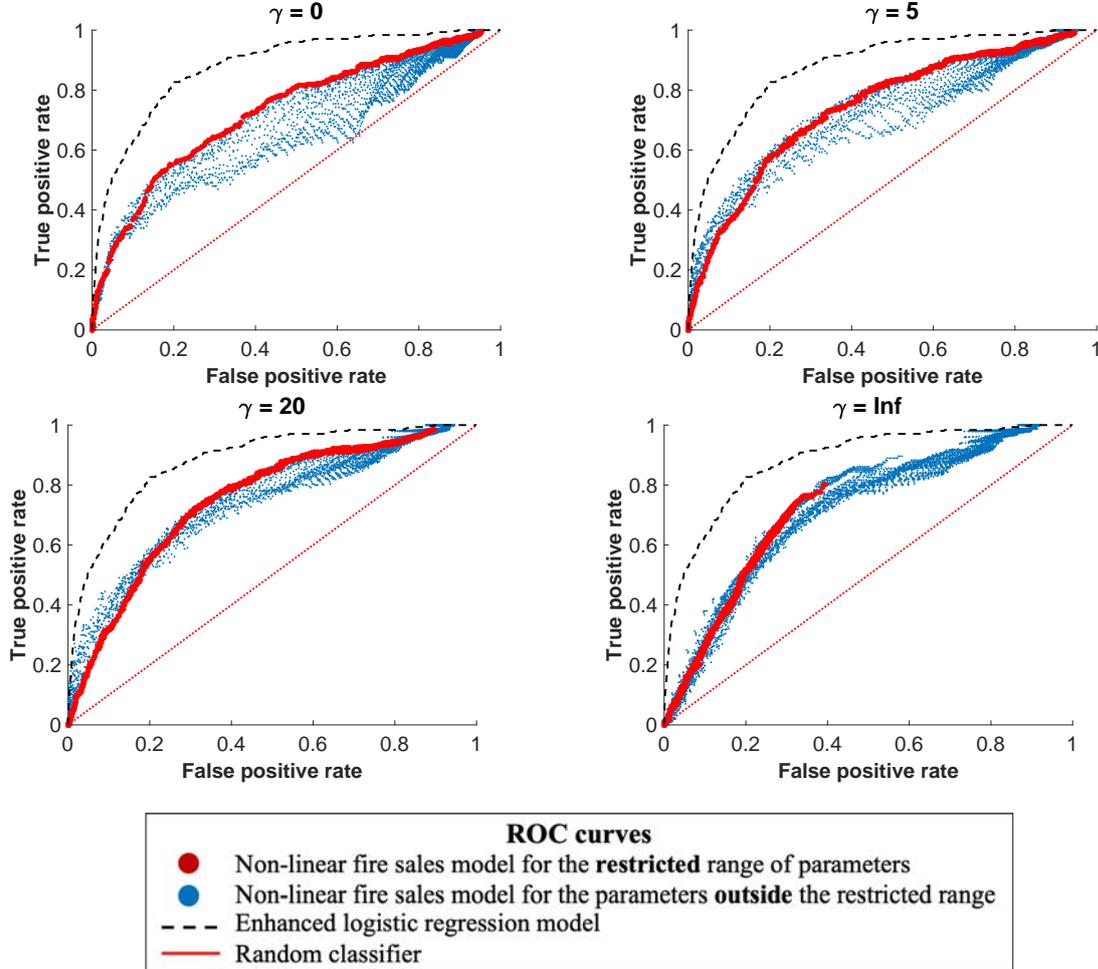


Fig. 9: **ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with initial shock on loans for construction and land development (asset class 5).** Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

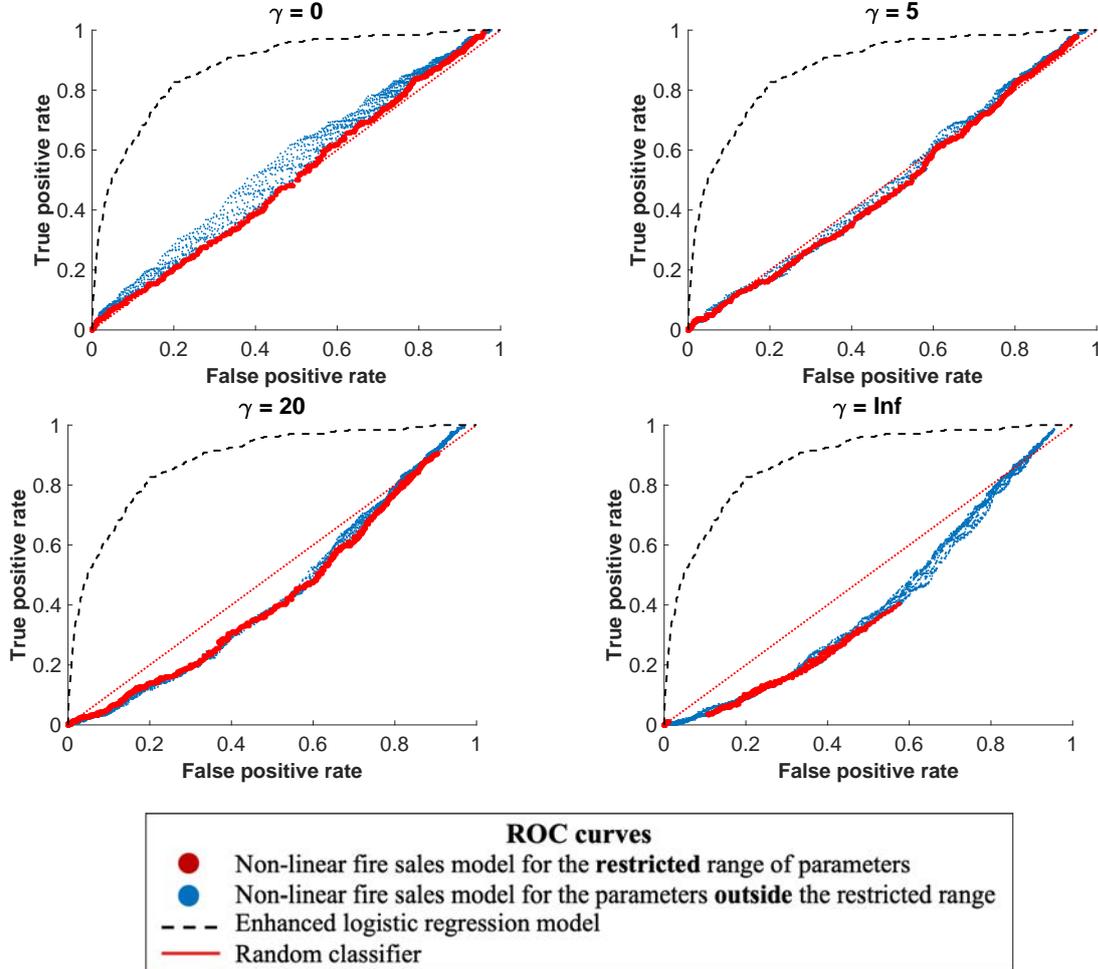


Fig. 10: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with initial shock on loans for construction and land development (asset class 8). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

4.4.1 Heterogeneity of market impact

So far, we considered the case of a homogeneous market impact parameter $\alpha_j = \alpha$ for all assets (except for cash). In the following, we relax this assumption and explore to what extent this improves the model performance. Our 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, we assume that the two relevant asset classes (1 and 5), are ten times less liquid compared to the other asset classes. [Figure 11](#) and [Figure 12](#) compare the model performances for homogeneous (top panels) versus heterogeneous market impact parameters (bottom panels) for the cases $\gamma = 0$ ([Figure 11](#)) and $\gamma = 20$ ([Figure 12](#)). Each column in the figures corresponds to an initial shock on a specific asset class. For all asset classes, we see that ROC curves for heterogeneous market impact are superior to those for homogeneous market impact.

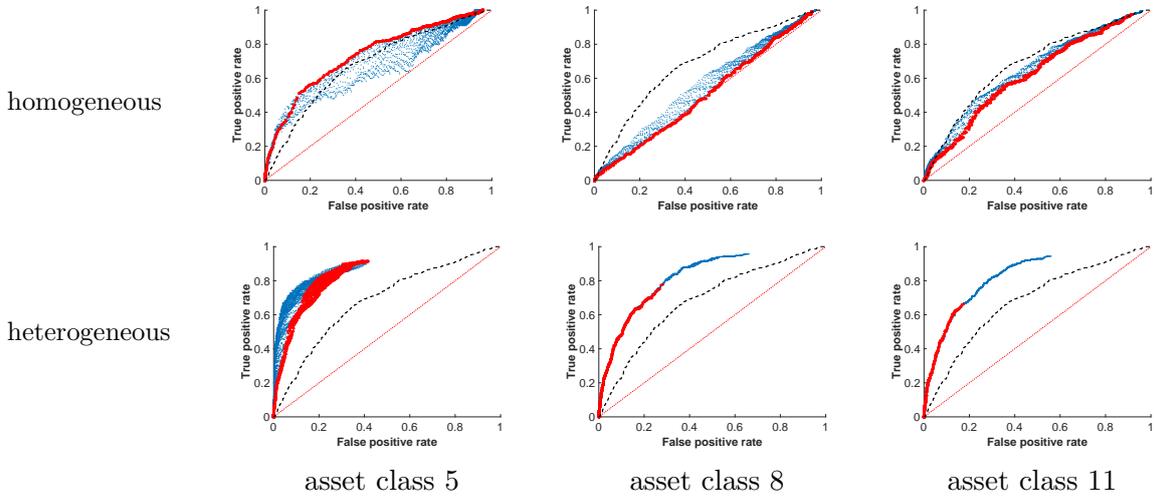


Fig. 11: Comparison between ROC curves of model with homogeneous market impact parameters (top panels) versus model with heterogeneous market impact (bottom panels). Each column corresponds to an initial shock on a specific asset class (see [Table 1](#) for the classification). Here we consider $\gamma = 0$.

Overall, the performance of the model improves for our 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 we 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).

4.4.2 The impact of multiple rounds of asset liquidations

So far we only included one round of asset liquidations in our stress test model. Here we explore to what extent accounting for multiple rounds affects the predictive performance of the model. In particular, we 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

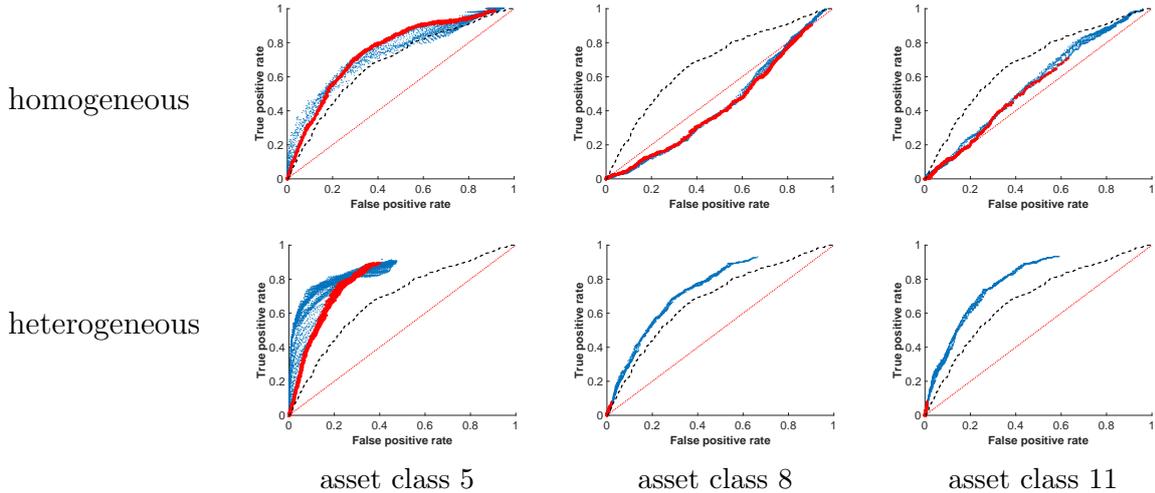


Fig. 12: Comparison between ROC curves of model with homogeneous market impact parameter (top panels) versus model with heterogeneous market impact parameter (bottom panels). Each column corresponds to an initial shock on a specific asset class (see Table 1 for the classification). Here we consider $\gamma = 20$.

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.

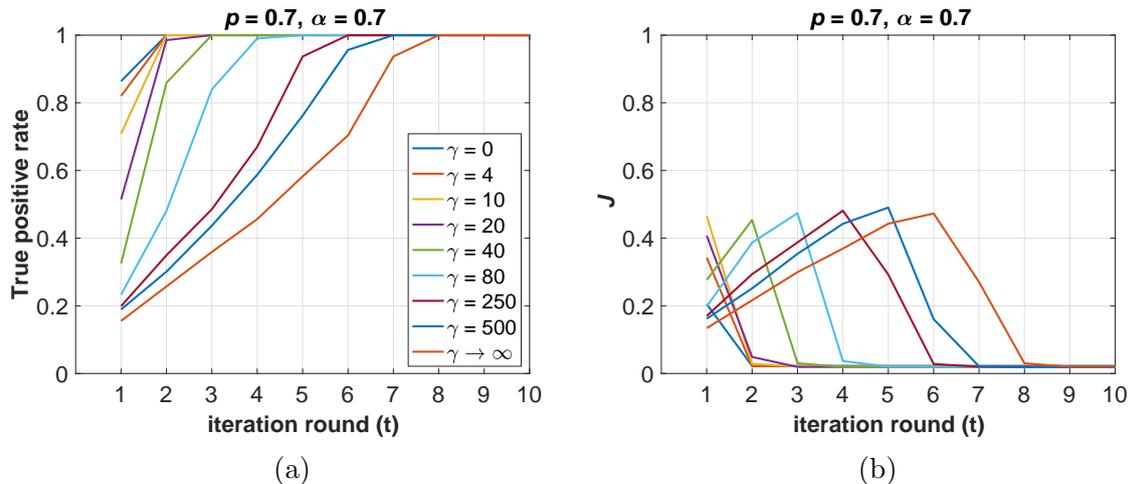


Fig. 13: (a) The number of bank failures that is correctly identified by the model (normalized to the actual number of bank failures), as a function of γ and iteration round (t). (b) The corresponding value of J as a function of γ and iteration round (t). The results correspond to a combination of p and α in the restricted parameter range ($p = 0.7$ and $\alpha = 0.7$).

We look at the effect of multiple rounds of liquidations along two different dimensions: first, we check whether it increases the TPR (the number of bank failures that the model correctly predicts) of the different models. To this end, we plot the TPR, for several values of γ , as a function of the iteration round (t) in Figure 13(a).¹² In line

¹²In the plot, we only show the results for $p = \alpha = 0.7$ that correspond to an example from within the restricted range of parameters that we considered ($p, \alpha \in [0.6, 1]$). We 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, we show the same plot for $p = \alpha = 0.5$.

with the results of [Cont and Schaanning \(2017\)](#), we 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 $\gamma \rightarrow \infty$ reaches $TPR = 1$ at $t = 8$.

Second, we investigate to what extent the increased TPR from [Figure 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, we plot J as a function of the iteration round (t) in [Figure 13\(b\)](#). Here we again observe that the dynamics of J depend on the assumed liquidation dynamics: for small values of γ (e.g. $\gamma = 0, 4, 10, 20$), we 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 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. We 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 test model.

5 Conclusion

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

Our 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 a number of assumptions, in particular with regards to behavioral response of banks. In this paper, we showed that it is important to consider a range of assumptions in relation to the behaviour of banks. We provided a framework to do this in a structured manner by means of a free model parameter (γ) and illustrated how the optimal 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, we 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 test model.

Our findings suggest several interesting paths for future research. First and foremost, we think it is of utmost importance to perform similar analyses on other data sets. This paper specifically focuses on the behavior 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 test model. Lastly, in line with most of the existing literature we focused on a single layer of the banking network. Future work should include additional layers in order to provide a more complete assessment of financial system vulnerabilities.

References

- Adrian, T. and H. S. Shin (2010). Liquidity and leverage. *Journal of Financial Intermediation* 19(3), 418–437.
- Baranova, Y., J. Coen, P. Lowe, J. Noss, and L. Silvestri (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*.
- Bardoscia, M., F. Caccioli, J. I. Perotti, G. Vivaldo, and G. Caldarelli (2016). Distress propagation in complex networks: the case of non-linear debtrank. *PLoS ONE* 11(10).
- Battiston, S., M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli (2012). DebtRank: too central to fail? financial networks, the fed and systemic risk. *Scientific Reports* 2(541).
- Caccioli, F., M. Shrestha, C. Moore, and J. D. Farmer (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance* 46, 233–245.
- Cavestany, R. and D. Rodríguez (2015). Backtesting, stress testing and sensitivity analysis. In R. Cavestany, B. Boulwood, and L. Escudero (Eds.), *Operational Risk Capital Models*. Risk Books.
- Cole, R. A. and L. J. White (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 E. Schaanning (2017). Fire sales, indirect contagion and systemic stress testing. *Norges Bank Working Paper 2*.
- Daniélsson, J. (2011). Backtesting and stress testing. In *Backtesting and Stress Testing*, pp. 143–166. West Sussex: John Wiley & Sons Ltd.
- Di Gangi, D., F. Lillo, and D. Pirino (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control* 94, 117–141.

- Drucker, S. and M. Puri (2009). On loan sales, loan contracting, and lending relationships. *Review of Financial Studies* 22(7), 2835–2872.
- Duarte, F. and T. M. Eisenbach (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*. New York: Academic Press.
- Eisenberg, L. and T. H. Noe (2001). Systemic risk in financial systems. *Management Science* 47(2), 236–249.
- Elsinger, H., A. Lehar, and M. Summer (2006, sep). Risk assessment for banking systems. *Management Science* 52(9), 1301–1314.
- Fricke, C. and D. Fricke (2020). Vulnerable asset management? the case of mutual funds. *Journal of Financial Stability Forthcoming*.
- Glasserman, P. and H. P. Young (2015). How likely is contagion in financial networks? *Journal of Banking and Finance* 50, 383–399.
- Glasserman, P. and H. P. Young (2016). Contagion in financial networks. *Journal of Economic Literature* 54(3), 779–831.
- Greenwood, R., A. Landier, and D. Thesmar (2015). Vulnerable banks. *Journal of Financial Economics* 115(3), 471–485.
- Grigat, D. and F. Caccioli (2017). Reverse stress testing interbank networks. *Scientific Reports* 7(15616).
- Haldane, A. G. and R. M. May (2011). Systemic risk in banking ecosystems. *Nature* 469, 351–355.
- Huang, X., I. Vodenska, S. Havlin, and H. E. Stanley (2013). Cascading failures in bi-partite graphs: model for systemic risk propagation. *Scientific Reports* 3(1219).
- Jiang, H., D. Li, and A. W. Wang (2017). Dynamic liquidity management by corporate bond mutual funds. Available at SSRN: <https://ssrn.com/abstract=2776829>.
- Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010). Did securitization lead to lax screening? evidence from subprime loans. *Quarterly Journal of Economics* 125(1), 307–362.
- Levy-Carciente, S., D. Y. Kenett, A. Avakian, H. E. Stanley, and S. Havlin (2015). Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance* 59, 164–181.
- Niepmann, F. and V. Stebunovs (2018). Modeling your stress away. *International Finance Discussion Papers* 1232(7).
- Nier, E., J. Yang, T. Yorulmazer, and A. Alentorn (2007). Network models and financial stability. *Journal of Economic Dynamics & Control* 31, 2033–2060.

- Philippon, T., P. Pessarossi, and B. Camara (2017). Backtesting european stress tests. *NBER Working Paper No. 23083*.
- Ramadiah, A., F. Caccioli, and D. Fricke (2020). Reconstructing and stress testing credit networks. *Journal of Economic Dynamics and Control* 111.
- Siemsen, T. and J. Vilsmeier (2018). On a quest for robustness: about model risk, randomness and discretion in credit risk stress tests. *Deutsche Bundesbank Discussion Paper* 31.
- Swets, J. A., R. M. Dawes, and J. Monahan (2000, oct). 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.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3(1), 32–35.

A Appendix

A.1 Classification of existing network models of contagion due to common asset holdings

In the main text, we consider network models of contagion due to common asset holdings, for which two classes of dynamics have been proposed in the literature: *threshold dynamics* (Caccioli et al., 2014; Huang et al., 2013), and *leverage targeting* (Greenwood et al., 2015). Here we 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 A.1, we 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). They 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	
		<i>linear</i>	<i>non-linear</i>
Leverage targeting	<i>not included</i>	Huang et al. (2013)	Caccioli et al. (2014)
	<i>included with threshold</i>		Cont and Schaanning (2017)
	<i>included</i>	Greenwood et al. (2015)	Caccioli et al. (2014) (extended)

Table A.1: Comparison between different existing network models of contagion due to common asset holdings, based on the type of market impact function used and whether leverage targeting is included or not.

A.2 Logistic regression

In the main text, we plot the ROC curve of a standard logistic regression model. This allows to make in-sample predictions by modelling the conditional probability of observing bank i 's default. We denote this as $Prob(y_i = 1|X = x_i)$, where x_i is a vector of leverage and total assets of bank i (both log transformed). Logistic regression specifies this probability as:

$$Prob(y_i = 1|X = x_i) = \frac{1}{1 + \exp(-x_i\beta)}, \quad (27)$$

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

Table A.2 shows the results. As it can be seen from the table, the explanatory variables (leverage and total assets) are significantly positively associated with the probability of default.

	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 A.2: Results from a logistic regression model to explain bank failures where the dependent variable takes a value of one if a bank failed, and zero otherwise. The explanatory variables are leverage and total assets (log transformed). *** indicates statistical significance at the 0.1% level.

A.3 Enhanced logistic regression

Table A.3 shows the results of the enhanced logistic regression model with additional explanatory variables.

	Coefficient	Standard error	<i>t</i>-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 A.3: 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.

A.4 Initial shock versus contagion dynamics

In the main text, we focus on bank failures identified from the contagion dynamics alone, and ignored failures due to the initial shock. In Figure A.1, we 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, we find that the results of the second category are superior compared to those we presented in the main text. In other words, the results shown in the main text likely underestimate the performance of the network approach.

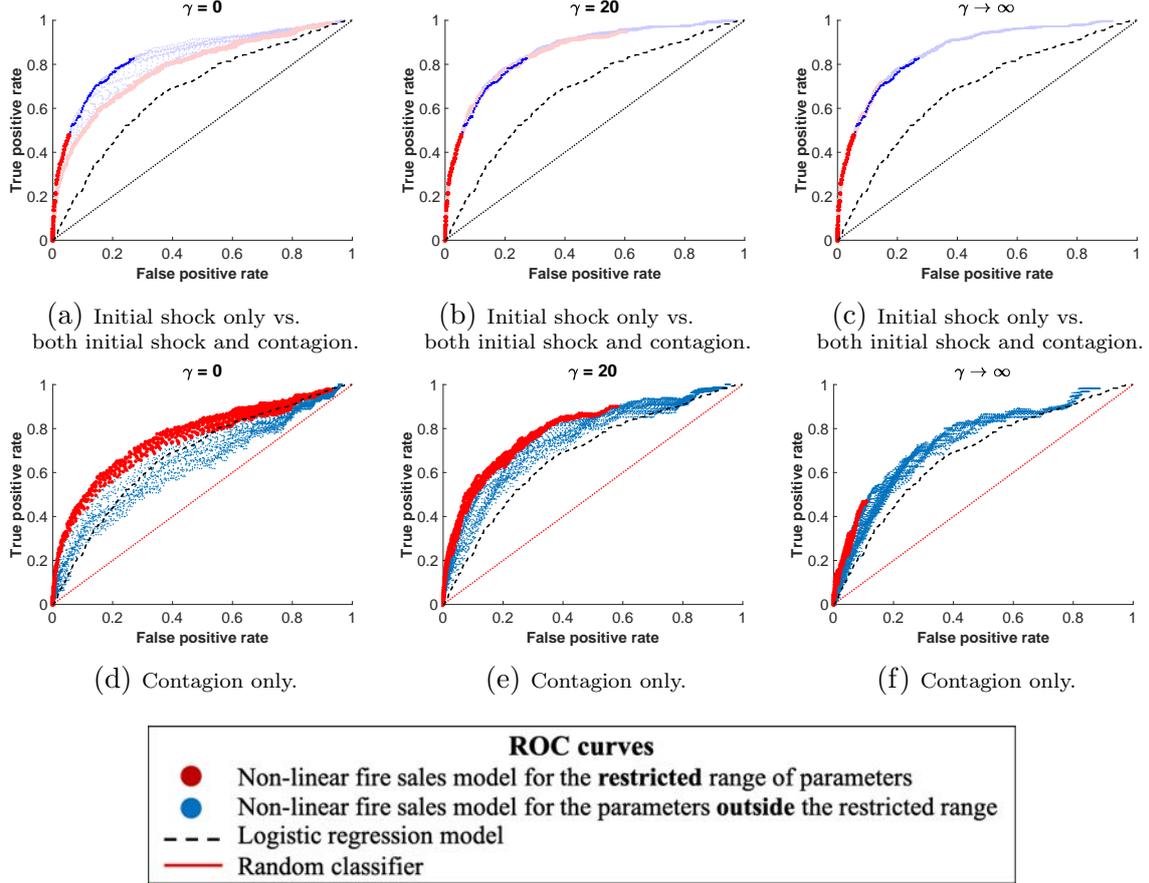


Fig. A.1: ROC curves of the defaulted banks predicted by the model with $\gamma = 0$ (left panels), $\gamma = 20$ (middle panels) and $\gamma \rightarrow \infty$ (right panels), with an initial shock on asset class 1 (loans for construction and land development). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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.

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 A.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.

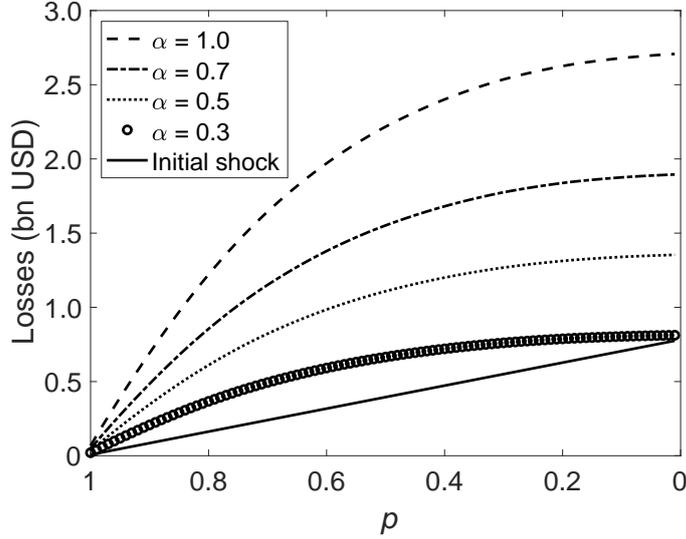


Fig. A.2: Total losses (in billion USD) due to the initial shock and due to the contagion dynamics only (first round of liquidation), for $\gamma = 0$. Results for initial shock on asset class 1 (loans for construction and land development). The total pre-shock assets in the network are 12.7 bn USD. 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.

A.5 Shocking less relevant asset classes

We found that two asset classes (1 and 5) were most relevant in terms of our model application. For the sake of completeness, here we show results when imposing an initial shock on asset 8 (loans to individuals). Figure A.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.

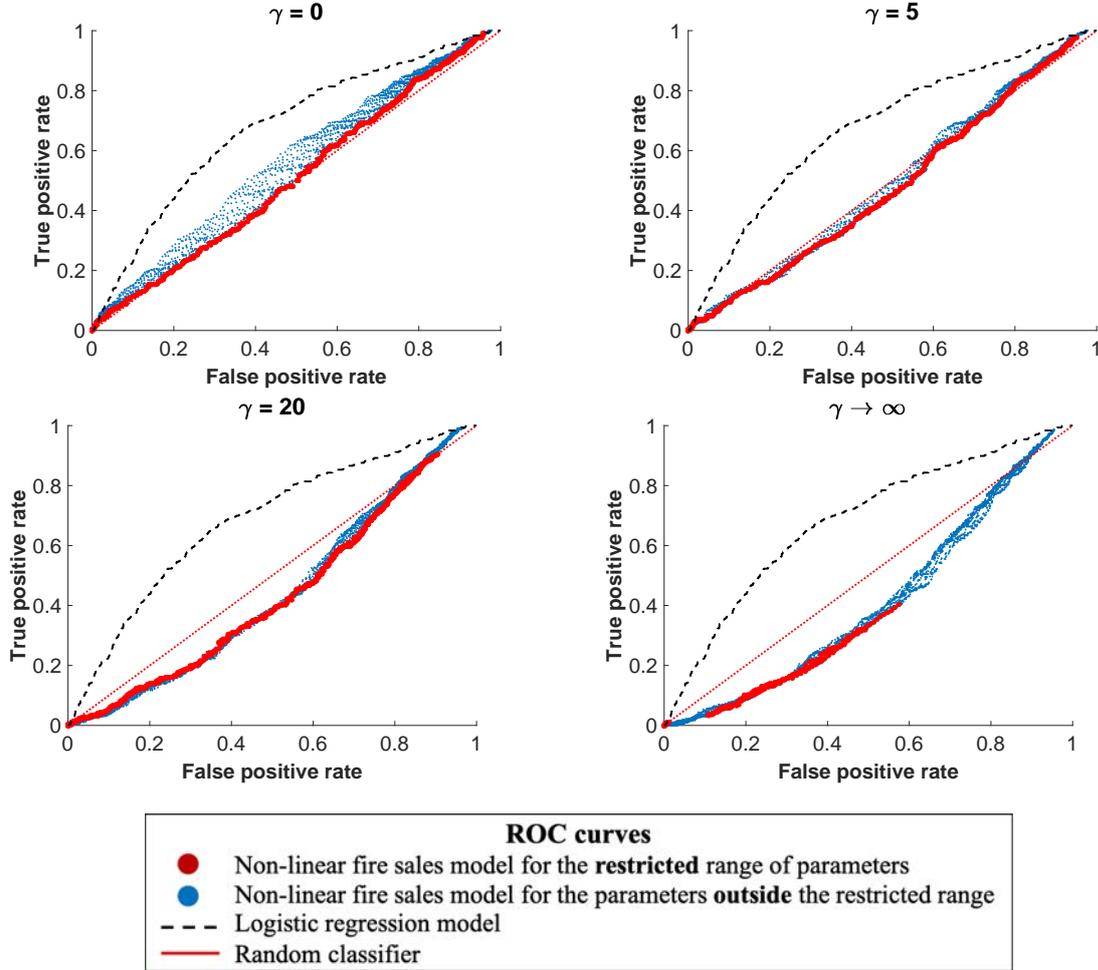


Fig. A.3: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with an initial shock on loans to individuals (asset class 8). Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

A.6 Multiple liquidation rounds outside the restricted parameter range

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

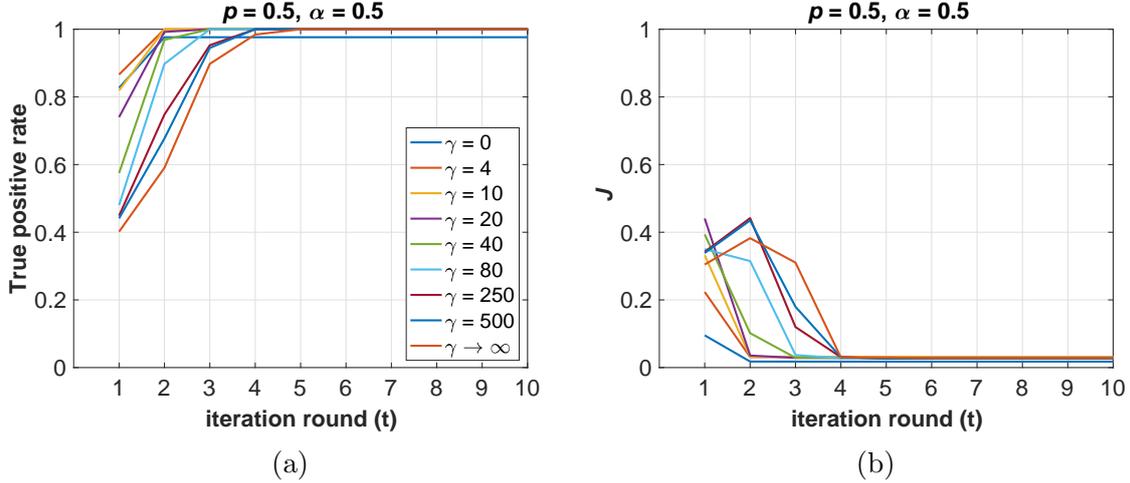


Fig. A.4: (a) The number of bank failures that is correctly identified by the model (normalized to the actual number of bank failures), as a function of γ and iteration round (t). (b) The corresponding value of J as a function of γ and iteration round (t). The results correspond to a combination of p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

A.7 Model accuracy with a non-linear market impact function

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

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

for a range of α as in the main text, $\alpha \in \{0, 0.01, 0.02, \dots, 0.99, 1\}$. We plot the corresponding ROC curves of this choice of Ψ_j in [Figure A.5](#) (for initial shock on asset: loans for construction and land development) and in [Figure A.6](#) (for initial shock on asset: loans secured by non-farm non-residential properties). We find that the results of linear and non-linear market impact are qualitatively similar (see [Figure 4](#) and [Figure 5](#) to compare with linear market impact).

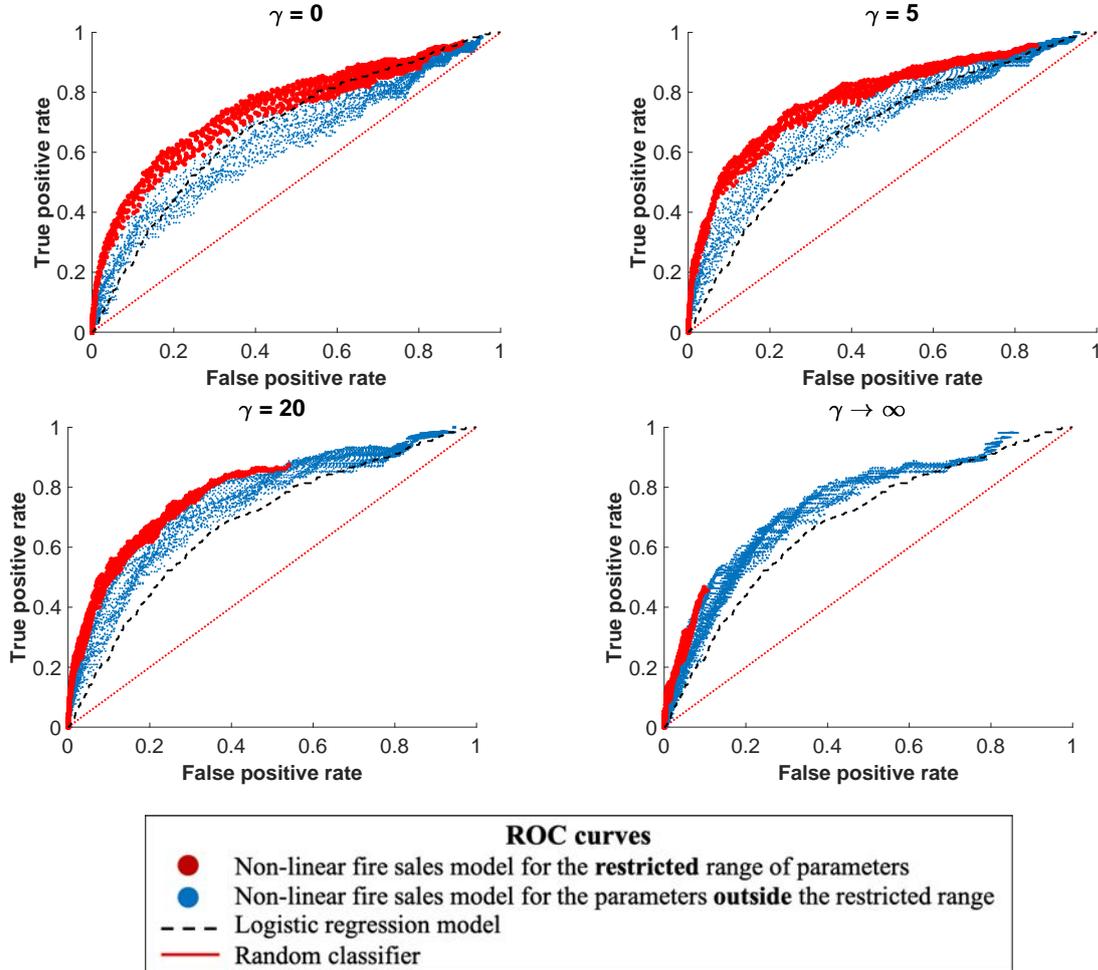


Fig. A.5: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with an initial shock on loans for construction and land development (asset class 1). The plot is similar to Figure 4, but here we consider a non-linear market impact function. Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.

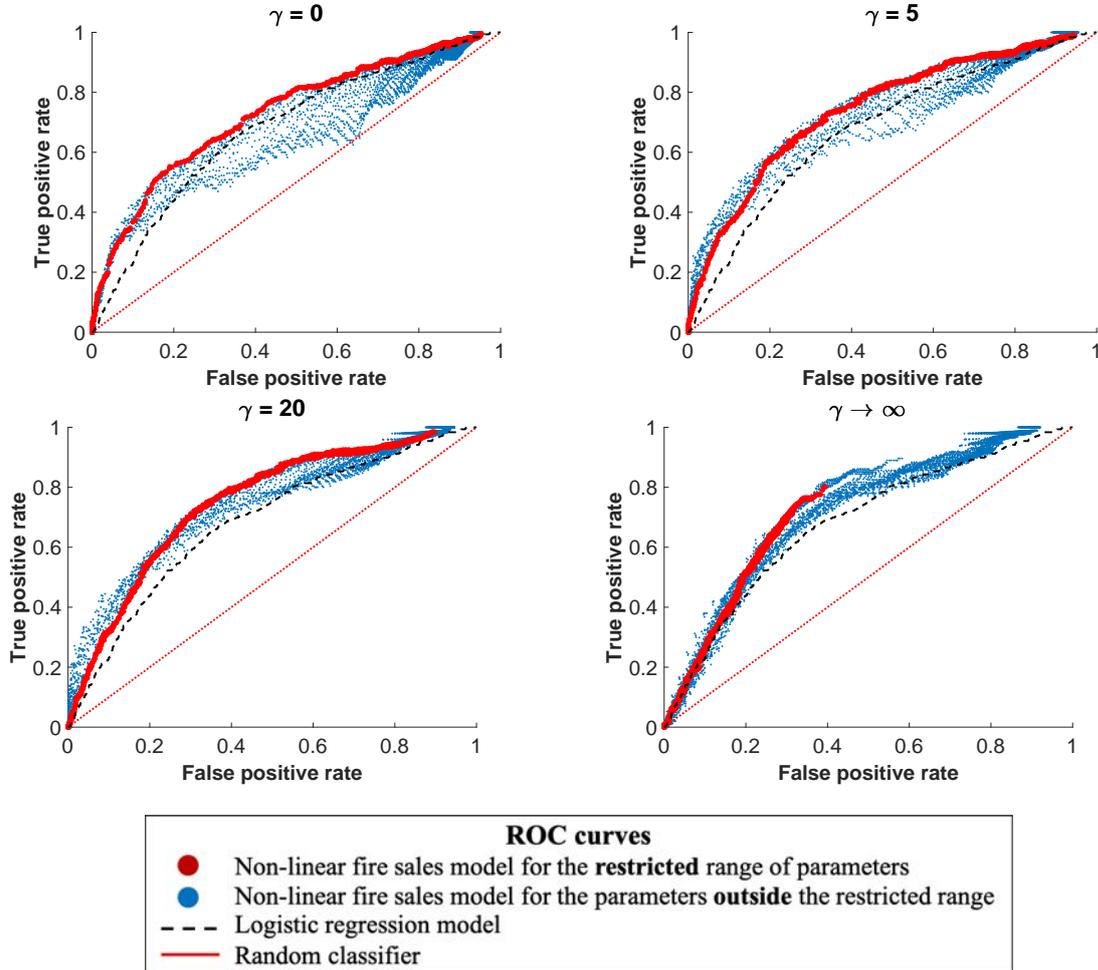


Fig. A.6: ROC curves of the model with $\gamma = 0, 5, 20$ and $\gamma \rightarrow \infty$, resp., with an initial shock on loans secured by non-farm non-residential properties (asset class 5). The plot is similar to Figure 5, but here we consider a non-linear market impact function. Each dot represents a true positive/false positive rate pair for a specific combination of the initial shock (p) and the market impact parameter (α). We 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 we consider only the first round of asset liquidations, and exclude bank failures due to the initial shock in the model assessment.