A Structural Model of Interbank Network Formation and Contagion*

Patrick Coen[†] Jamie Coen[‡]

April 2025

We study the equilibrium relationship between interbank exposures and bank default risk: how exposures affect risk, and how banks account for this when forming the exposures network. We leverage novel data on aggregate interbank exposures across multiple types of financial instrument. We find that contagion is material (risk would be 0.2% higher if exposures increased by 1%) but that banks account for this in equilibrium (risk would be 10% higher if they did not). We also find systematic heterogeneity in contagion based on the characteristics of the banks involved, with implications for the identification of systemically important banks and regulation.

Keywords: Contagion, systemic risk, interbank network, network formation.
JEL Codes: G21, G28, G01, L14.

*The views in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are particularly grateful to Alessandro Gavazza and Christian Julliard for many helpful discussions. We are also grateful for comments by Celso Brunetti (discussant), Áureo de Paula, Anil Kashyap, Péter Kondor, Caterina Lepore, Simon Mayer (discussant), Martin Oehmke, Claudia Robles-Garcia, Rhiannon Sowerbutts, Yinan Su (discussant), John Sutton, Dimitri Vayanos, Mungo Wilson, Tanju Yorulmazer (discussant), Shengxing Zhang and Jean-Pierre Zigrand, as well as seminar participants at the Bank of England, the Federal Reserve Bank of New York, Imperial College, the London School of Economics, Princeton, Stanford GSB, the Toulouse School of Economics, Universidad Pompeu Fabra, Queen Mary University of London, and conference participants at the BSE Summer Forum: Networks, RES Junior Symposium, EARIE, the Econometric Society Winter Meeting, FMCG, FIRS, the SAMMF OTC Workshop and the TSE-Banque de France Workshop. We are grateful to the Bank of England for providing the data. Both authors acknowledge the financial support of the Economic and Social Research Council. Patrick Coen also acknowledges the financial support the French National Research Agency under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010) and of the TSE-Partnership foundation, a research organization under the aegis of Toulouse School of Economics, sponsored by several partners (the list of which is available at www.tse-fr.eu/tsepartnership).

†Toulouse School of Economics, patrick.coen@tse-fr.eu.
‡Imperial College London and Bank of England, w.coen@imperial.ac.uk.

1 Introduction

Banks have direct exposures to each other, brought about by transactions across a broad range of financial instruments. While these exposures exist for a reason and must create some gains to trade, they also create counterparty risk. There are numerous examples throughout history of exposures held by banks creating contagion, bank failures, and in some cases systemic financial crises (Basel Committee, 2014).

Unsurprisingly, contagion between banks has been the subject of significant regulatory scrutiny (Yellen, 2013) and academic attention (Glasserman and Young, 2016). There is a large theoretical literature on contagion (Acemoglu et al., 2015), as well as empirical papers that model interconnections between banks in specific instruments and markets (Craig and Ma, 2022; Gofman, 2017; Li et al., 2022). However, there is little empirical work studying aggregate exposures across instruments and markets, primarily because of the absence of data. In this paper, we fill this gap using novel data from the Bank of England in which major banks periodically reported their aggregate exposures to their most important counterparties.

We use these novel data to do two things. First, we document various facts on the size and variation in exposures between banks, as well as their relationship with bank risk. Second, we structurally estimate a model of how the exposures network is formed in equilibrium and how it affects bank risk. Based on these we make the following contributions. First, we quantify the joint determination of risk and exposures, including the equilibrium elasticity of risk to exposures and of exposures to risk. Second, we document heterogeneity in these effects: some exposures involve systematically lower contagion than others based on the characteristics of the banks involved. Third, we unpack what this means for policy, including the identification of systemically important banks and the effect of regulatory caps on large exposures.

The starting point for our work is Bank of England data on interbank exposures. These data are collected by the Bank of England through periodic regulatory surveys of 18 of the largest global banks from 2011 to 2018, in which they report the exposures they have to their most important banking counterparties. The data are novel, relative to the data commonly used in this literature, in two ways that are important for our context: (1) the data include a broad range of instruments, making them the best available proxy for a bank's aggregate exposure to another bank and (2) the data contain rich detail on the types and characteristics of the instruments that make up each exposure.

We document the variation in this empirical network along various dimensions. We find that exposures are large: a single bank's aggregate exposures to the other 17 banks in our sample can exceed one third of its equity. The network is also dense: we find non-zero exposures between over two-thirds of bank-pairs. However, these links are heterogeneous: there is wide variation in the size of exposures across bank-pairs and in the instruments that form them. Finally, we look for systematic variation in this heterogeneity, and in particular co-movement with default risk, for which we proxy using credit default swap spreads. We find that the exposure of bank i to bank j is positively correlated with the default risk of both parties. However, when we instrument for default risk with a measure of bank fundamentals that comes from banks' heterogeneous loadings on regional equity indices, we find that this correlation turns negative.

Based on these facts, we build a model of network formation and contagion. The primary feature of the model is that banks incur a funding cost when they take on an exposure, and that this funding cost is increasing in their default risk. We model default risk as spatially autocorrelated through the exposures network, a workhorse empirical model in network econometrics (De Paula, 2017) and finance (Denbee et al., 2021) that represents contagion as a network spillover. Together, these assumptions rationalize our reduced-form findings in an intuitive way: banks trade off the surplus they gain from an exposure against the resulting increase in risk, and thus in funding cost. Exposures are positively correlated with risk because of contagion, but negatively correlated with fundamentals because of the way that banks manage this contagion.

As well as enabling quantification and counterfactual analysis, the model plays an important role in measurement. We show how banks take into account risk spillovers when forming the network, meaning that we can infer information about these spillovers based on their observed exposures. This enables us to measure contagion with more power and additional heterogeneity, even while conditioning on a broad range of fixed effects. We structurally estimate both the network formation and contagion parts of the model, based on the joint empirical distribution of exposures, risk and bank fundamentals.

We use this estimated model to make three contributions. First, we quantify the joint determination of exposures and risk in equilibrium. We find that a 1% increase in all bank exposures would increase mean bank default risk by 0.23%%: there is, in other words, material contagion. Banks, however, respond endogenously to this contagion by scaling back their exposures to risky banks: we find that a 1% increase in bank risk would decrease mean bank exposures by 0.64%. In a counterfactual simulation we turn off this endogenous response to contagion by banks, and find that exposures would be 6% larger and mean bank risk would be 10% higher. We believe this to be the first quantification in the literature of the joint determination of aggregate exposures and risk.

Second, we uncover heterogeneity in contagion in the cross-section: some exposures involve systematically lower contagion than others based on the characteristics of the banks involved. We find robust evidence that contagion is lower when the exposure is held by a bank with larger equity and when the banks involved have similar fundamentals. We find weaker evidence that contagion is higher when the exposure is related to derivative rather than securities financing or fixed income exposures. Based on these contagion shifters, we find substantial heterogeneity in contagion across banks and pairs of banks: the largest spillover across bank-pairs is 7 times the mean.

Third, we show what our results imply for policy. We show that the systematic heterogeneity we find in contagion matters for identifying systemically important banks, where such banks face stricter regulation (Basel Committee, 2014). There are various measures of systemic importance (or, equivalently in our context, network centrality), but in general terms a bank is deemed systemically important if it has large exposures to other systemically important banks. If, however, spillovers are heterogeneous and exposures are formed endogenously, then this ranking is likely to be biased: *some exposures are large because their spillovers are small.* In other words, banks with central positions in the network might have such positions because their spillovers, and thus their contribution to systemic risk, are low. We propose an alternative measure of systemic importance based on network data that is weighted by heterogeneous spillovers. This weighted centrality measure implies materially different centrality rankings among banks: the bank that is the second most systemically important in our sample based on the unweighted network is only the ninth most important based on our alternative weighted centrality measure.

We then consider the implications of our findings for the regulatory cap on individual exposures (Basel Committee, 2014, 2018b). We counterfactually simulate such a cap, as well as an alternative cap on the aggregate exposures held by a given bank across all its counterparties. We find that the aggregate cap is preferable to the individual cap, in the sense that for any individual cap one can design an aggregate cap that achieves both lower risk and higher surplus. Bilateral caps potentially penalize links that are large precisely because they are safe, and also allow for leakage, whereby banks can respond to such caps by increasing their smaller, uncapped exposures. Aggregate caps avoid these two features, and thus can better manage the risk-reward trade-off in these markets.

We discuss related literature below. In Section 2, we describe our data and set out our reduced-form findings. In Section 3, we set out our model. In Section 4, we describe empirical approach. In Section 5, we report our results. In Section 6, we conclude.

1.1 Related literature

Our work is related to three strands of literature: (i) the effects and formation of financial networks, (ii) structural estimation of financial markets and (iii) optimal regulation in financial markets.

There is an extensive literature on the effect of network structure on outcomes in financial markets, both theoretical (Acemoglu et al., 2015; Ballester et al., 2006; Eisfeldt et al., 2023; Elliott et al., 2021, 2014; Galeotti and Ghiglino, 2021; Gofman, 2011; Nier et al., 2007) and empirical (Denbee et al., 2021; Eisfeldt et al., 2023; Iyer and Peydro, 2011). These spillovers can be direct or indirect, such as informational spillovers along the network (Acharya and Yorulmazer, 2008; Goldsmith-Pinkham and Yorulmazer, 2010). There is also a large theoretical literature on network formation in financial markets (Acharya and Bisin, 2014; Babus, 2016; Cabrales et al., 2017; Chang and Zhang, 2018; Farboodi, 2023; Shu, 2022), and a smaller empirical literature (Blasques et al., 2018; Cohen-Cole et al., 2010; De Paula, 2020; Elliott et al., 2021; Ellul and Kim, 2021).

Our paper builds on detailed, granular data from the Bank of England on aggregate exposures between global banks, which we use to estimate a model in which exposures result in contagion, which banks take into account when they form networks. Langfield et al. (2014) describes the setting and various empirical facts using the same data. Aldasoro and Alves (2018) study similar data for European banks only. Many network formation papers explain sparse endogenous networks through fixed costs to bilateral link formation (De Paula, 2020). Our setting is quite different, in that the empirical network in aggregate exposures that we seek to match is dense, but heterogeneous in the size of individual exposures. We rationalize this with heterogeneity in the marginal cost of holding an exposure, given its effect on risk. Our empirical setting allows us to estimate rich heterogeneity in network spillovers, which we show has important implications and regulation. Our contribution is to quantify the joint determination of exposures and bank default risk.

There is a related literature on structural estimation of trade between banks in financial markets, including Allen and Wittwer (2023), Brancaccio and Kang (2022), Coen and Coen

(2022) and Pinter and Uslu (2022) in search-based models of financial markets. Benetton (2021) and Benetton et al. (2025) consider the effects of regulation in mortgage markets. There are three particularly important papers on bank lending and systemic risk in financial networks: Craig and Ma (2022) estimate a model of interbank lending in Germany based on monitoring costs, Li et al. (2022) estimate a model of deposits and the US bank payments network, and Gofman (2017) estimates a random network formation model for the US Fed Funds market. Our contribution is to focus on aggregate exposures and their direct effect on bank risk, using the identifying information contained in both CDS-implied bank default risk and the observed network.

We also contribute to the literature regarding optimal regulation in financial markets (Baker and Wurgler, 2015; Batiz-Zuk et al., 2016; Duffie, 2017; Greenwood et al., 2017). We directly model the impact of regulation on both aggregate interbank surplus and bank default risk. Our work is part of a growing effort in the literature to take a more holistic view of global banks across their multiple activities (Benetton et al., 2022; Diamond et al., 2024). We remain largely agnostic in the broader debate over whether banks should be regulated more tightly (Admati, 2016; Hellmann et al., 2000), by focusing on the *design* of regulation given heterogeneity across banks, rather than on the *level* of regulation.

2 Data and Facts

We first describe our data. We then use these data to document novel empirical facts about total bank exposures, and to discuss their implications for the remainder of this paper.

2.1 Data

We define in general terms the exposure of bank i to bank j at time t as the immediate loss that i would bear if j were to default, as estimated at time t and assuming a recovery rate of zero. The way in which this is calculated varies from instrument to instrument, but in general terms this can be thought of as (1) the value of the instrument, (2) less collateral, (3) less any regulatory adjustments intended to represent variations in value or collateral in case of default (for example, regulation typically requires a "haircut" to collateral when calculating exposures, as in the event of default assets provided as collateral are likely to be worth less). The procedure to calculate these exposures is determined by regulation (Basel Committee, 2018b). We set out more detail on the data in the Internet Appendix.

We use regulatory data on bilateral interbank exposures, collected by the Bank of England. The dataset offers a unique combination of breadth and detail in measuring exposures. Much of the existing literature (such as Denbee et al. (2021)) on empirical banking networks relies on data from payment systems. This is only a small portion of the activities that banks undertake with each other and is unlikely to adequately reflect the extent of interbank activity or the risk this entails.

18 of the largest global banks operating in the UK report their top 20 exposures to banks over the period 2011 to 2018. Banks in our sample report their exposures every six months from 2011 to 2014, and quarterly thereafter. They report exposures across fixed income instruments, securities financing transactions – including repo, reverse repo, and securities lending and borrowing – and derivatives. The data are censored: we only see each bank's top 20 exposures, and only if they exceed £10 million. The data include granular breakdowns of each exposure: by type (e.g. they break down derivatives into interest rate derivatives, credit derivatives etc.), currency, maturity and, where relevant, seniority.

We use this dataset to construct a series of snapshots of the interbank network between these 18 banks. We calculate the total exposure of bank i to bank j at time t, which we denote C_{ijt}, as the sum of exposures across all types of instrument in our sample. The result is a panel of $N = 18$ banks over $T = 21$ periods from 2011 to 2018 Q2, resulting in $N(N-1)T = 6,426$ observations.

Although the dataset includes most of the world's largest banks, it omits banks that do not have a subsidiary in the UK.[1] Our sample of banks includes 5 UK banks and 13 subsidiaries of banks with headquarters outside the UK. For the non-UK banks that are included in our dataset, we observe only the exposures of the local sub-unit, and not the group. For non-European banks, this sub-unit is typically the European trading business. We consider this partial observation directly in Section 4.

2.2 Empirical facts

The data has two important advantages over much of the data used in the literature. First, our dataset is the closest available representation of *total* exposures, when most other empirical assessments of interbank connections rely on a single instrument, such as CDS (Eisfeldt et al., 2023) or overnight loans (Denbee et al., 2021). Second, our data are on exposures, rather than simply market value, in that when banks report their exposures they account

[1] This is particularly relevant for some major European investment banks, who operate branches rather than subsidiaries in the UK, and hence do not appear in our dataset.

for collateral and regulatory adjustments. Data based solely on market value is arguably a representation of bank activity, rather than counterparty risk. We describe our treatment of the data in more detail in the Internet Appendix. See Langfield et al. (2014) for a detailed overview of the data.

These data advantages allow us to set out three empirical facts about how total exposures vary along various dimensions. First, we document substantial cross-sectional and inter-temporal variation in the size of exposures between banks. Second, we document systematic variation and show how exposures co-vary with bank default risk, including the risk of the bank holding the exposure and the risk of the bank to which the exposure is held. Third, we show there is also variation at a more granular level in the types of instrument that make up these total exposures. We discuss each fact in turn, before considering their implications and how they guide our work in the rest of this paper.

Fact 1: Exposures are large and heterogeneous

First, we document the intensive margin of exposures, and show that there is substantial variation in the intensity of network links as measured by the size of the exposure. Figure 1 shows the network of exposures between banks in H1 2015. In Table 1, we report various percentiles of the distribution of exposures by year. Both Figure 1 and Table 1 show there is substantial cross-sectional heterogeneity: the 95th percentile of the exposures across all years is 14 times the median. There is also inter-temporal variation: the mean exposure in 2011 is over 50% larger than the mean in 2016. This inter-temporal variation is not smooth or obviously related to macro-economic variation.

Table 1 also makes clear that the exposures are large. The largest exposure in our sample is £7,682m, the largest total exposures to other banks in a given period is £26,367m. The mean exposure is £285m and the mean total exposure to other banks in a given period is £4,850m. In aggregate, these exposures are large both in absolute terms and as a percentage of banks' equity. The average aggregate exposure for a bank is almost £5bn and represents 9% of a bank's equity, with large variation across banks. Together, these facts imply that risk on these exposures has the potential to significantly impact banks' overall risk.

Turning to the extensive margin, our sample is effectively limited to the core of the banking network and does not include its periphery. Our observed network is, therefore, dense: of the $N(N-1)T$ possible links in our sample, only approximately 30% are 0.

Fact 2: Exposures co-move with bank default risk

Having established in fact 1 that there is significant variation in exposures, we now look for sources of systematic variation. In particular, in this paper we ultimately seek to study the joint distribution of bank default risk and exposures: how risk responds to exposures through contagion, and how banks take that into account when forming exposures. In this empirical fact, we set out reduced-form facts on this joint distribution.

To proxy for bank default risk, we use the spreads on publicly traded credit default swaps (obtained from Bloomberg), which as shown by Hull et al. (2009) and Allen et al. (2021) can be used to compute the (risk-neutral) probability of default. Let p_{mt} be this measure of default risk for bank m at time t, measured in basis points. We summarize the variation in this variable across banks and through time in the first two columns of Table 2. There is significant heterogeneity across banks, and variation through time, with CDS premia decreasing through our sample.

In the first two columns of Table 3 we report the results of a regression of C_{ijt} on p_{it} and p_{jt}, along with various fixed effects. We find that exposures and bank default risk co-move positively. This result is mechanically consistent with contagion: if default risk is increasing in bank exposures because of counterparty risk, then it is unsurprising that exposures and default risk co-move positively. To isolate the response of exposures to default risk, rather than the reverse, we need variation in default risk that is unrelated to bank exposures.

To obtain such variation, we construct a panel of 'fundamentals' for banks through time, using a shift-share approach. We begin by collecting quarterly data for 7 regional equity indices over the period 2004 to 2018.[2] We show variation through these time in Figure 2. We compute the return for index k at time t as the simple average of the index's return at t relative to one quarter before and one year before. We use these averaged returns to ensure our fundamentals capture both short-term and longer-term variation. We regress each bank's default risk p_{it} on this set of returns R_{kt} during a training period before our sample from 2004 to 2011, with bank-specific coefficients.

This procedure gives us a set of loadings for each bank on each equity index, estimated before the beginning of our sample. We then construct bank fundamentals during our sample period by multiplying these bank-specific loadings by realized returns on the regional equity indices during our sample period. We refer to this as a bank's fundamentals and denote it

[2]The indices are for the UK, Europe, Japan, Latin America, the US, China and Hong Kong, and are provided by S&P. We plot these indices and show their variation in Figure 2.

in the remainder of this paper by X_{it}.

We summarize this variable in the final two columns of Table 2. There is significant variation both through time and across banks, which will be key when it comes to estimating our model in Section 4. In Table 4 we show the results of regressing risk p_{it} on fundamentals X_{it} and various fixed effects over our sample period. The two variables are closely correlated, with a deterioration in our measure of fundamentals correlated with increased risk.

This approach is predicated on banks having different geographic footprints and as a result different sensitivities to regional macroeconomic conditions. A macroeconomic shock in Japan, for example, affects US, European and Asian banks differently depending on the nature and structure of their business. We thus use such regional shocks throughout our sample period as a set of shocks that shift bank default risk differentially.

In the final two columns of Table 3 we report the results of a regression of C_{ijt} on p_{it} and p_{jt}, instrumented by these bank fundamentals X_{it} and X_{jt}. Continuing the preceding example, instead of studying how exposures to and from a Japanese bank respond to its default risk, we study how they respond to shocks to a Japanese equity index. We find that the signs reverse relative to the OLS regression: C_{ijt} is decreasing in both the risk of bank i and bank j. That is, riskier banks receive less and supply less exposures. The effects are economically meaningful: a one standard deviation increase in the risk of the j bank is associated with a £85mn decrease in individual exposures, equivalent to 30% of the mean.

This is consistent with banks accounting for risk when they determine the exposures network. In the model that follows we suggest a mechanism for how and why they do this. The response of exposures to bank fundamentals is also consistent with the behavior of the overnight interbank network during the 2008 crisis. First, risky banks were not supplied; in other words, they experienced lockout (Welfens, 2011), consistent with our finding that C_{ijt} is lower when bank j's fundamentals are worse. Second, risky banks did not supply (Gale and Yorulmazer, 2013), consistent with our finding that C_{ijt} is lower when bank i's fundamentals are worse. Third, in the worst periods of the financial crisis there was effectively market shutdown in markets for certain instruments, in that trade effectively ceased (Afonso et al., 2011; Allen et al., 2009), when the fundamentals of all banks were poor. We discuss the implications of these results in more detail below.

Fact 3: Characteristics

We observe granular data on the type of instruments that make up these exposures, which we summarize in Table 5. For each exposure C_{ijt}, we compute the share that relates to derivatives, securities financing transactions, and fixed income exposures, and summarize the distribution of these variables. Around 60% of exposures relate to derivatives, with a further quarter relating to repo and securities lending, and the remainder representing holdings of other banks' debt. We observe the residual maturity of banks' repo and securities lending and borrowing, as well as the maturity of their debt holdings. Much of this activity is short term, with around 70% of less than 3 months' residual maturity.

The contracts that make up interbank activity vary both through time and across banks. We report the standard deviation in these characteristics through time and across banks, and show that cross-sectional variation is more important than inter-temporal variation. In other words, banks seem to consistently supply the same types of exposure. We consider in estimation below whether this heterogeneity in exposure characteristics matters for contagion.

2.3 Implications

We draw three implications from these facts. First, interbank exposures are large, both individually and in aggregate. Second, they are heterogeneous, both in their size and their characteristics. Third, there is systematic variation between exposures, bank risk and bank fundamentals that is consistent with contagion.

These facts leave the following open questions. How big or small is contagion? How does it impact bank risk and the network of exposures between banks? How does it vary across heterogeneous banks and links? What implications does this have for policy? To answer these, we build and estimate a model.

3 Model

We first introduce the setup of the model and notation. We then describe the default risk process and the network formation game. Our objective is to focus on the fundamental tradeoff between risk propagation and surplus creation, but in a way that is tractable, consistent with our empirical facts (and in particular the substantial heterogeneity in exposures across banks) and estimable. We do this with a model that is analytically simple, but that contains rich heterogeneity that we can estimate given our granular data.

3.1 Setup and notation

There are N banks. At time t, the interbank network consists of an $N \times N$ directed adjacency matrix of total exposures, $\mathbf{C_t}$. C_{ijt} is the element in row i and column j of $\mathbf{C_t}$, and indicates the total exposure of bank i to bank j at time t. $\mathbf{C_t}$ is directed in that it is not symmetric: bank i can have an exposure to bank j, and bank j can have a (different) exposure to bank i. $\mathbf{p_t}$ is an $N \times 1$ vector of bank default risks: the element in position i is the probability of default of bank i. $\mathbf{p_t}$ is a function of $\mathbf{C_t}$ and an $N \times K$ matrix of exogenous bank fundamentals, which we denote $\mathbf{X_t}$. This function is the default risk process, and the effect of $\mathbf{C_t}$ on $\mathbf{p_t}$ represents contagion, as we will define more formally below.

C_{ijt} results in profits to bank i (we term this supply of exposures) and to bank j (demand for exposures). These profits depend on bank default risk via the funding costs of banks, but not via uncertainty in the returns to holding an exposure, in a way we will formalize below. This allows us to include a role for bank default risk while retaining the tractability needed to take our model to data. Banks choose their supply and demand decisions simultaneously. For simplicity, there is no friction between changes in bank fundamentals and the formation of the network: once fundamentals change, the equilibrium network changes immediately, such that our model is focused on long-run effects.

3.2 Default risk process

By the default risk process, we mean the process by which bank fundamentals and the network of exposures combine to result in bank default risk. We model a bank's default risk process as a spatially autocorrelated regression, as is commonly used in network econometrics (De Paula, 2017) and specifically in finance (Herskovic et al., 2020; Ozdagli and Weber, 2017):

$$\underbrace{\mathbf{p_t}}_{\substack{\text{Default}\\\text{risk}}} = \underbrace{\mathbf{X_t}\boldsymbol{\beta}}_{\substack{\text{Funda-}\\\text{mentals}}} + \underbrace{(\boldsymbol{\Gamma}_\mathbf{t} \circ \mathbf{C_t})\mathbf{p_t}}_{\substack{\text{Counterparty}\\\text{risk}}} + \epsilon^{\mathbf{p}}_{\mathbf{t}} \tag{1}$$

where default risk vector $\mathbf{p_t}$, network matrix $\mathbf{C_t}$ and fundamentals vector $\mathbf{X_t}$ are defined in the previous sub-section, $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients on bank fundamentals, $\boldsymbol{\Gamma}_\mathbf{t} > \mathbf{0}$ is an $N \times N$ matrix of parameters that determine the effect of exposures on default risk through counterparty risk, $\epsilon^{\mathbf{p}}_{\mathbf{t}}$ is a $N \times 1$ vector of shocks and $\circ$ signifies the Hadamard product.

The only departure from a standard spatially autocorrelated regression is that the parameter governing the size of the network spillover (counterparty risk in our context), Γ_{ijt}, is allowed to be heterogeneous across bank pairs and time. $\Gamma_{ikt} > \Gamma_{imt}$ implies that $\frac{\partial p_{it}}{\partial p_{kt}} > \frac{\partial p_{it}}{\partial p_{mt}}$ for any common $C_{ikt} = C_{imt}$. That is, bank i's default risk is more sensitive to exposures to bank k than to bank m, holding exposures and fundamentals constant.

There are various reasons why this spillover could be heterogeneous. In principle, two things determine the impact on the default risk of bank i of a given exposure to a given counterparty j: (1) the recovery of that exposure in the event of default by j and (2) bank i's financial state in the event of default by j. This is in the spirit of distance-to-default models (Merton, 1974). This could support broad heterogeneity in the cross-section or time-series in the effect of a given exposure on bank i, depending on the characteristics of the banks involved. We do not embed these factors in our model, but instead leave Γ_{ijt} general in the spatially autocorrelated relationship implied by Equation 1. In our empirical analysis below, we parameterize Γ_{ijt} as a function of bank and bank-pair characteristics that proxy for these factors.

The partial equilibrium net effect of an exposure C_{ijt} is as follows:

$$\frac{\partial p_{it}}{\partial C_{ijt}} = -\Gamma_{ijt} p_{jt}$$

An exposure C_{ijt} has a greater effect on the default risk of bank i if the counterparty is particularly risky (p_{jt} is large) or the link is particularly risky (Γ_{ijt} is large).

To find equilibrium default risk we solve for a fixed point in $\mathbf{p_t}$. Subject to standard regularity conditions on $\mathbf{\Gamma}$ and $\mathbf{C}$ this spatially autocorrelated process can be inverted and expanded as follows:

$$\mathbf{p_t} = (\mathbf{I} - \mathbf{\Gamma} \circ \mathbf{C_t})^{-1}(\mathbf{X_t}\boldsymbol{\beta} + \mathbf{e^p_t}) = \sum_{s=0}^{\infty}(\mathbf{\Gamma} \circ \mathbf{C_t})^s(\mathbf{X_t}\boldsymbol{\beta} + \mathbf{e^p_t}) \tag{2}$$

3.3 Network formation

We model the supply of and demand for exposures, which collectively lead to network formation. As described above, there is material heterogeneity in the size and type of exposure across time. This has implications for the specificity with which we model the payoffs to demanding financial products, in that we cannot use a standard model of, for example, debt or CDS exposures. Instead, we use linear demand as a simple reduced-form that is applicable across a range of financial instruments but that allows significant heterogeneity when we take the model to data. On the supply-side, the empirical network we are seeking to model is dense with heterogeneous intensities. This leads us to focus on heterogeneity in *marginal* cost, rather than on fixed costs to link formation.

Bank i supplies exposures C_{ijt} and demands C_{kit}. On the supply-side, bank i incurs a cost from supplying and in return receives an interest rate from the demanding bank:

$$\Pi_{it}^{S} = \sum_{j} C_{ijt}(r_{ijt} - puc_{it})$$

where puc_{it} is i's per-unit-cost of an exposure and r_{ijt} is the interest rate paid by j to i. We formalize the key mechanism in the model in Assumption 1 below, in which we relate the bank cost function to its default risk.

Assumption 1. The effect of bank risk on bank cost: *We assume that the per-unit cost to bank i of holding an exposure is increasing in the default risk of bank i and its capital requirement:*

$$puc_{it} = l_{it}\, \phi\, p_{it}$$

where l_{it} is the capital requirement applied to bank i's exposures at time t and ϕ is a parameter governing the intensity of the relationship between risk and cost. This cost could be thought of as a funding cost, which is increasing in the bank's risk and in the amount of equity that it is required to raise. This assumption links the default risk process with network formation: banks care about their default risk because it affects their costs. We discuss this assumption, and its purpose, further below. Risk matters only for the banks' costs, as there is no uncertainty in the returns to holding an exposure. We formalize the way in which banks take this into account in the following assumption.

Assumption 2. Banks take the risk of other banks as given: *Bank i internalizes the direct effect of its network choices on p_{it}, but takes all other bank default risks $p_{k\neq i,t}$ as given.*

We think Assumption 2 is a reasonable representation of how banks' risk management and capital allocation processes work. Banking regulation and supervision are designed to force banks to consider the impact of the exposures they create on their risk.[3] This

[3] See, for example, https://www.bis.org/basel_framework/chapter/CRE/20.htm.

regulation in principle takes into account the risk of a bank's counterparty, but takes this risk as given. Bajaj et al. (2018) provide evidence that banks' capital allocation decisions are largely based on these regulatory frameworks, and as a result are also likely to take other banks' risks as given. Assumptions 1 and 2 together imply that there are cost externalities across the network, in that whilst banks consider the direct effect of their actions on their own default risks, they do not directly account for the fact that their choices may make other banks riskier and lead them to face higher costs.

Turning to the payoff that j receives when it is supplied exposures, we model it simply as a quadratic in $\mathbf{C_t}$:

$$\Pi_{jt}^{D} = \sum_i \zeta_{ijt} C_{ijt} - \frac{1}{2}\sum_i C_{ijt}^2 - \sum_i r_{ijt} C_{ijt}$$

where ζ_{ijt} represents heterogeneity in the sensitivity of the bank j's technology to product i. In effect, this is a reduced-form in which the gains to trade are left fixed and encoded in ζ_{ijt}.[4]

The problem of bank i is to choose its exposure supply $\{C_{ijt}\}_j$ and exposure demand $\{C_{kit}\}_k$ to maximize its profits, taking $p_{k\neq i}$ and all interest rates as given, and subject to non-negativity constraint $C_{ijt} \geq 0$ and $C_{kit} \geq 0$:

$$\begin{aligned} max_{\{C_{ijt}\}_j, \{C_{kit}\}_k} \quad & \sum_k \zeta_{kit} C_{kit} - \frac{1}{2}\sum_k C_{kit}^2 - \sum_k r_{kit} C_{kit} \\ & + \sum_j C_{ijt}(r_{ijt} - l_{it}\, \phi\, p_{it}(\mathbf{C_t})) \end{aligned}$$

where we emphasize that bank i's risk and thus funding cost are functions of the network $\mathbf{C_t}$, following Equation 2 and Assumption 1. Taking the first order condition for bank i demand from bank k, C_{kit}, and re-arranging:

$$r_{kit} = \zeta_{kit} - C_{kit} \tag{3}$$

[4]Note that we could include a non-monetary payoff to bank i from supplying the exposure C_{ijt}, analogous to technology parameter ζ_{ijt} on the demand-side. We choose not to do this because separate demand- and supply-side parameters could not be identified from the network data, and in any case all that matters for our analysis is the aggregate returns to trade, not how those returns are split between the two parties.

In other words, our functional form assumptions imply that the bank demanding exposures has linear inverse demand with heterogeneous intercepts depending on technology ζ_{kit}. Taking the first order condition for bank i supply to bank j, C_{ijt}, and re-arranging:

$$r_{ijt} = \underbrace{l_{it}\phi p_{it}}_{\text{Per-unit cost}} + \underbrace{\Gamma_{ijt}p_{jt}\sum_{m} C_{imt}l_{it}\phi}_{\text{Change in total cost}} \tag{4}$$

Bank i, when choosing to supply C_{ijt}, therefore balances the return it gets from supplying against the effect of its supply on its marginal cost, where this marginal cost depends on its per-unit cost and the marginal change in its total cost through the effect of its supply on its default risk, via the default risk process described above.

3.4 Equilibrium

Definition 1. Equilibrium: *We define a competitive equilibrium in each period t as an $N \times N$ matrix of exposures $\mathbf{C_t}$ and $N \times 1$ vector of default risks $\mathbf{p_t}$ such that markets clear and every bank chooses its exposures optimally given the actions of other banks.*

For interior solutions where $C_{ijt} > 0$, market clearing simply equates Equations 3 and 4 for the supply and demand of C_{ijt}, which when re-arranged implies:

$$C_{ijt} = \zeta_{ijt} - l_{it}\phi p_{it} - \Gamma_{ijt}p_{jt}\sum_{m} C_{imt}l_{it}\phi \tag{5}$$

This implies that C_{ijt} depends on the need of j for i's product, the cost to i of supplying, and the effect of the transaction on the risk (and thus cost) of both parties. This condition is linear in $\mathbf{C_t}$ taking $\mathbf{p_t}$ as given, but nonlinear in $\mathbf{C_t}$ taking into account its effect on $\mathbf{p_t}$ through the default risk process. Equilibrium $\mathbf{C_t}$ are non-linear in bank fundamentals, given that $\mathbf{C_t}$ and $\mathbf{p_t}$ enter multiplicatively. This Equation 5, together with Equation 2 of the default risk process, pins down equilibrium network exposures.

3.4.1 Solving for an equilibrium

Each bank chooses their exposures optimally, given non-negativity constraints, such that individual equilibrium exposures can be 0 or strictly positive. This implies the usual system of first order conditions and complementary slackness conditions based on Equation 5. If one substitutes Equation 2 into these optimality conditions, then this gives a system of equations in $\mathbf{C_t}$. It is a system of infinite-length series of polynomials, such that in general no analytical solution exists. Instead, we solve these equilibrium conditions numerically. In practice, we use the following algorithm:

1. Start with an initial guess of $\mathbf{C_t}$ in which all of its elements are equal to 0.
2. Given this guess of $\mathbf{C_t}$, calculate $\mathbf{p_t}$ using Equation 2.
3. Given this $\mathbf{p_t}$, calculate $\mathbf{C_t}$ using Equation 5 and complementary slackness conditions.
4. Repeat steps 2 and 3 until convergence.

This replaces a full non-linear search over $\mathbf{C_t}$ with a series of linear inversions, so it is quick and reliable, and performs consistently well for any reasonable parameterization.

We make no general claims about uniqueness or existence, but confirm using numerical simulation that at our estimated results the resulting equilibrium network is unique. We test this by fixing all parameters at their estimated values, for which we know an equilibrium $\mathbf{C_t}$ using our algorithm above. We then numerically search for other equilibria that satisfy the full-set of non-linear equilibrium conditions, from 100 sets of randomly chosen starting points around the known equilibrium. In each case, the numerical solver converges to the single known equilibrium.[5]

3.5 Characterizing the equilibrium and discussion

Having described the model, we now characterize the equilibrium and in doing so discuss why this model suits our research question and data.

First, the model proposes a mechanism for the joint determination of exposures and default risk: banks obtain some surplus from their exposures, but these exposures lead to counterparty risk, which in turn leads their funding costs to increase. This rationalizes why, as set out in our empirical facts above, exposures are sensitive to the risk of the bank holding the exposure and the risk of the bank in which the exposure is held. The model has an explicit role for regulation as part of this mechanism, and it is one that speaks to the

[5] We choose starting values by drawing from a normal distribution with mean 0 and standard deviation equal to the standard deviation of the known equilibrium $\mathbf{C_t}$. We add this random noise to the known equilibrium to get a set of starting values, from which we then numerically search for equilibrium exposures that satisfy the optimality conditions. In Table 6 we report the mean and maximum absolute difference between (i) the starting values and the known equilibrium and (ii) the chosen numerical solution and the known equilibrium. We show that despite meaningful variation in the starting values, all of the numerical optima are very close to the known equilibrium.

main objection to capital regulation voiced by banks: that it makes their operations more costly.

Second, there is a simple and intuitive market failure: banks care about their own default risk and cost, but they do not take into account the effect of their choices on the default risk of other banks. The result is that equilibrium exposures may be too large and banks too risky, consistent with existing work on network externalities (Acemoglu et al., 2015) and, given that we translate this into bank cost, cost externalities across firms.

Third, the model is consistent with anecdotal evidence of what happened during the financial crisis: the relationship between risk and bank cost means that risky banks are less likely to supply exposures and to be supplied exposures. In extreme cases, the combined effects of these channels mean the equilibrium network could be empty. The model is also consistent with our empirical facts. There are various sources of heterogeneity across banks, pairs of banks and time, including bank fundamentals X_{it}, bank technology or demand for particular exposures ζ_{ijt} and, in the case where we estimate heterogeneous spillovers, Γ_{ijt}.

Fourth, the model is simple but has rich parameterized heterogeneity. Building up from its simplest version, suppose $\phi = \Gamma_{ijt} = \zeta_{ijt} = 0$ for all i, j, t. In this case ("autarky"), there are no network links in equilibrium and risk p_{it} is driven solely by exogenous fundamentals. If $\zeta_{ijt} > 0$ ("exogenous network and risk"), then equilibrium network links become positive but exogenous. If $\phi = 0$, $\Gamma_{ijt} > 0$ ("exogenous network and endogenous risk"), then risk responds to banks' network formation, but banks have no reason to account for it. If $\phi > 0$, $\Gamma_{ijt} > 0$ ("endogenous network and risk"), then equilibrium risk responds to the equilibrium network, and vice versa because of the effect of risk on cost. This parameterization includes an important role for heterogeneous network spillovers Γ_{ijt} in the cross-section or time-series, which induce variation in the marginal costs of risk formation: exposures between pairs of banks for which spillovers are high, for example, are more costly.

The trade-off between surplus generation and contagion and cost is governed by the size of the surplus (technology ζ_{ijt}), the effect of exposures on risk (spillovers Γ_{ijt}) and the effect of risk on cost (cost parameter ϕ). Finally, the size of the inefficiency is mediated by the size of the higher-order network spillovers that banks neglect in their network formation, governed by the size of the spillovers Γ_{ijt}.

Finally, the model provides primarily linear estimating equations such that this heterogeneity can be feasibly estimated. This allows us to benefit from the strengths of our data and, in particular, to tractably include a broad range of fixed effects.

4 Empirical approach

We first describe two parameterizations we make to take our model to data. We then describe the structure of our baseline estimation and show identification. We consider robustness to our baseline approach in the results section below.

4.1 Parameterizations

We make two parameterizations to take our model to data.

4.1.1 Heterogeneous contagion

As set out in Section 3, we allow for the network spillover term Γ_{ijt} to be heterogeneous in our model. We unpack this in estimation in various ways. First, we estimate a baseline in which network spillovers are homogeneous. Second, we estimate bank-varying spillovers terms in the i and j dimensions: $\Gamma_{ijt} = \Gamma_i$ and $\Gamma_{ijt} = \Gamma_j$, to show that there is empirical variation in the cross-section.

We then look to unpack the sources of this heterogeneity, by parameterizing Γ_{ijt} as a function of variables that may drive contagion. In principle, two things matter when considering the impact on the default risk of bank i of a given exposure to a given counterparty j: (1) the recovery of that exposure in the event of default by j and (2) bank i's financial state in the event of default by j. This is in the spirit of distance-to-default models (Merton, 1974), for example. Our objective is to identify observable variables that might proxy for these two things, and then to test whether they matter for contagion. We call such variables contagion shifters, and incorporate them into the model using a linear specification censored at 0:

$$\Gamma_{ijt} = max(\sum_q \tilde{\Gamma}_q g^q_{ijt}, 0) \tag{6}$$

where g^q_{ijt} is a contagion shifter for pair i, j at time t and $\tilde{\Gamma}_q$ is the loading on that shifter. We allow for spillovers to be heterogeneous in the following contagion shifters. See Table 7 for summary statistics on their variation.

First, we consider the equity of firm i at time t, as a direct proxy for its financial health. If this is large, then the effect of a given exposure on p_{it} should be low, as even if j defaults i will be far from default. This suggests a negative contagion loading for equity.

Second, we consider the degree of overlap in bank i and bank j's fundamentals. As described in Section 2.2, we regress each bank's default risk on various regional equity indices with bank-specific coefficients to represent bank fundamentals. These loadings give us a natural way of assessing the similarity of bank fundamentals across pairs of banks. We calculate the cosine similarity of these bank-specific coefficients for each pair, effectively measuring the extent to which they face similar shocks. A higher value for this variable indicates that banks have more closely related underlying fundamentals. The sign on this variable could be positive or negative. Dissimilar banks may benefit from risk sharing, such that when bank j is in distress bank i is not, and so the risk to i of an exposure to j the exposures is not great. On the other hand, bank i may expect to recover more of its exposure C_{ijt} if j is closely related. For example, a bank facing a loss to a bank that operates in the same jurisdiction may be prioritized by regulators in bankruptcy, as was the case in Iceland in the crisis.[6] Alternatively, banks may be better able to recognize and mitigate increased risks to similar banks than to fundamentally different banks.

Third, we study the role of the instruments that make up an exposure, by including the average share of exposures from bank i to bank j that are (a) derivatives and (b) securities financing transactions, with fixed income exposures the reference category. These exposures can differ in complexity, seniority in case of default (Antinolfi et al., 2015; Bolton and Oehmke, 2015) and maturity, all of which could matter for the recovery of exposures in the event of default.

Fourth, we study the role of overall financial volatility, using the VIX. Here, the sign is once again ambiguous. If broader financial volatility takes a safe bank i and puts it in a region where it is at risk of default, then i's exposures become riskier in response to this volatility. By contrast, if rising volatility means that the risk on bank i's other assets far exceeds the risk from its exposure to j, then bank i's risk would become less sensitive to j's risk.

Finally, we include an intercept as a contagion shifter in Equation 6. This means that our specification nests homogeneous network spillovers $\Gamma_{ijt} = \Gamma$, when all other contagion loadings are set equal to 0.

4.1.2 Partially observed network data

The second parameterization we make relates to the structure of our data, and in particular the fact that, as described in Section 2, for exposures held by non-British banks we only

[6] 'Icesave dispute resurrected in court', Financial Times, 2014. `https://www.ft.com/barrier/corporate/a120fbd8-b877-4a4a-a2a7-dfadff2036da`

observe local-unit-to-group exposures, under-estimating their total exposure. For exposures held by British banks, including their exposures to non-British banks, we observe total group-to-group exposures. For example, let *US1*, *US2* and *UK* represent anonymized banks, each of which has a *Global Group* and various *Local Units*. We observe the exposures of *UK Global Group* to the global groups of all other banks (total aggregate exposures, in other words). We observe the exposures of *US1 Local Unit* to *US2 Global Group* and *UK1 Global Group*.

Our model and question concern global exposures, which our unadjusted network data underestimates for non-British banks. To infer these global exposures, we make use of the fact that our data is complete along certain dimensions, but incomplete along others. Let C^{LU}_{ijt} represent the exposures held by the local unit of bank i, and C_{ijt} represents total exposures as previously. We assume that $C_{ijt} = C^{LU}_{ijt}\alpha_{it}$, where α_{it} is a bank-time specific scalar that effectively scales up local unit exposures to obtain global group exposures.

We then estimate α_{it} outside of our main estimation, by imposing symmetry in ingoing and outgoing exposures with respect to British banks. That is, we assume that total British exposures to *US1* (which we observe) are the same as total exposures held by *US1* to British banks (which we know up to scalar α_{it}). We interrogate our estimated α_{it} in our results below, as well as considering a robustness check in which we estimate our model without any adjustment of this sort to our network data.

4.2 Estimation overview

Applying our parameterizations to Equation 2 and 5 gives us the following estimation equations for interior network exposures and default risk:

$$C_{ijt} = -\phi \sum_q \tilde{\Gamma}_q g^q_{ijt} p_{jt} \sum_m C_{imt} l_{it} + FE_{it} + \epsilon^C_{ijt} \tag{7}$$

$$p_{it} = X_{it}\beta + \sum_j \sum_q \tilde{\Gamma}_q g^q_{ijt} C_{ijt} p_{jt} + FE_t + \epsilon^p_{it} \tag{8}$$

Latin characters C_{ijt}, p_{it}, l_{it}, g^q_{ijt} and X_{it} represent observable data that are inputs into our estimation. Greek characters ϕ, $\tilde{\Gamma}_q$ and β represent unknown coefficients that we are seeking to estimate, and which we collectively denote $\mathbf{\Theta}$. ϵ^C_{ijt} represents unobserved demand shocks that are not captured by the fixed effects and ϵ^p_{it} represents unobserved bank fundamentals. Without loss of generality we normalize C_{ijt} and p_{it} by their largest single value during our sample. We do not attempt to fit exposures that are 0, and for which the first order condition need not hold.

We include various fixed effects. In the default risk equations, we include time fixed effects to account for common variation in, among other things, the risk premium. In the network exposure equations, we include it fixed effects. This allows us to identify the network spillover terms $\tilde{\Gamma}_q$ solely by looking at how a given bank i in a given time period t changes its exposures in response to the risk of its counterparty j. To identify ϕ with additional power, we then decompose these estimated fixed effects into the following:

$$FE_{it} = \bar{\zeta} - \phi l_{it} p_{it} + \epsilon_{it} \tag{9}$$

where $\bar{\zeta}$ is some common technology parameter, $\phi l_{it} p_{it}$ represents bank i's funding costs and ϵ_{it} is some idiosyncratic technology need at the it level. This approach effectively allows us to identify the key network spillover terms with a demanding fixed effect, which we then unpack to help identify ϕ. In robustness checks, we also show the effect of adding more onerous fixed effects, namely at the ij and jt level.

Estimation is then standard. We use Generalized Method of Moments, where we calculate the errors in these equations given a parameter guess, $\boldsymbol{\epsilon}(\mathbf{C_t}, \mathbf{p_t}; \boldsymbol{\Theta})$, and interact them with a set of instruments, $\mathbf{Z_t}$, which we describe below, and which imply the set of moment conditions $\mathbb{E}[\mathbf{Z_t'}\boldsymbol{\epsilon}(\mathbf{C_t}, \mathbf{p_t}; \boldsymbol{\Theta})]$. We search numerically over parameters to match the empirical analogues to these moments as closely as possible. We calculate the standard gradient-based sandwich estimator for the asymptotic variance of $\boldsymbol{\Theta}$ assuming White heteroscedasticity (Greene, 2018), and weight the moments efficiently in a two-step GMM procedure.

4.3 Identification

We now turn to exactly which instruments go into the matrix $\mathbf{Z_t}$. We first set out instruments for the exposures equations and then for the default risk equations. We list all inputs to the structural estimation, including instruments, in Table 7.

The endogenous terms in Equation 7 are related to $p_{jt} \sum_m C_{imt} l_{it}$. An obvious shifter for this is $X_{jt} l_{it}$, the fundamental of j at time t based on its loadings on regional equity indices, as described in Section 2, interacted with capital regulation. In broad terms, if C_{ijt} is sensitive to the fundamentals of bank j then the model implies that network spillovers must be significant. Conversely, if C_{ijt} does not change when X_{jt} changes, then the model implies that network spillovers are small.

The model implies that $p_{jt} \sum_m C_{imt}$ is actually a non-linear function of X_{jt}, X_{it} and the fundamentals of other banks. We are thus not limited to using only $X_{jt}l_{it}$ as a shifter, but also choose a quadratic in X_{it} and X_{jt}, consisting of $X_{jt}l_{it}$, $X_{jt}^2 l_{it}$ and $X_{jt}X_{it}l_{it}$ (noting that terms involving only X_{it} would be swept up by FE_{it}). This closely mirrors the setting in Kelejian (1971), who proxies for an unknown reduced form for non-linear endogenous regressors using a polynomial in available instruments. We interact these three base instruments with FE_i to make use of the network structure of our data, and to account for the fact that heterogeneity in spillovers and technology means that the relationship between exposures and fundamentals need not be homogenous across banks. This results in $3N = 54$ instruments overall. We report our first stage results in Table 8.

We instrument for p_{it} in Equation 9 in an analogous way, using two instruments $l_{it}X_{it}$ and $l_{it}X_{it}^2$. That is, we conclude that ϕ is large if the average exposures held by i at time t are sensitive to its fundamentals and to capital regulation.

Finally, we also construct instruments in the default risk Equation 8 in an analogous way. The endogenous variables in Equation 8 take the form $C_{ijt}p_{jt}$. p_{jt} is a function of ϵ_{it}^p by construction due to network spillovers, for which Kelejian and Prucha (1998) and Kelejian and Prucha (1999) instrument in a standard spatial autoregression using fundamentals interacted with the exogenous network. In our setting, however, the network is endogenous. The advantage of explicitly considering network formation is that this gives us natural instruments for this endogenous network. As for network formation, above, we instrument for these endogenous terms with a quadratic in bank i's fundamentals and the fundamentals of other banks, again based on Kelejian (1971). We instrument for $\sum_j g_{ijt}^q C_{ijt}p_{jt}$ with X_{it}^2, $\sum_j g_{ijt}^q X_{jt}$ and $\sum_j g_{ijt}^q X_{jt}^2$ for each contagion shifter apart from the common intercept. This results in $2N^q - 1 = 9$ instruments overall, where $N^q = 5$ is the number of contagion shifters.

5 Results

We first set out our parameter estimates for models with increasingly rich heterogeneity in network spillovers. We then quantify the implications of our estimates for the joint determination of exposures and default risk in equilibrium. Having established that contagion is material and heterogeneous, and affects both network formation and contagion in equilibrium, we derive the implications for policy.

5.1 Contagion: magnitudes, heterogeneity and drivers

Table 9 reports out baseline estimates of our model parameters. In the first column we report estimates for our simplest specification, where the impact of an exposure C_{ijt} on the risk of firm i is constant across pairs and time. Our estimate of contagion Γ is positive and significant, indicating that an exposure between i and j causes bank i's risk to be increasing in bank j's risk.

We also find a positive and significant estimate of $\phi = 5.435$, which determines the cost to a bank of an exposures. A bank with a mean value of default risk and capital regulation – reported in Table 7 – faces a per-unit cost equal to 5.3%. This is reassuring, and supports the interpretation of this cost as a funding cost. Variation in risk causes significant variation in this cost across banks: a 1 standard deviation increase in risk leads to a 3% increase in costs.

We now study heterogeneity in contagion, a key focus of this paper. In the second and third column we allow contagion intensity to vary across i and j banks respectively. We report the average and standard deviation of these parameters across banks. In each case, we find strong evidence that (a) contagion is material on average, and (b) there is significant variation in contagion across banks.

Having established that contagion is material and heterogeneous, we now study its determinants. To do so, we parameterize contagion as a function of the contagion shifters we introduce in Section 4.1.1. We do this for an increasingly rich set of contagion shifters, and report the results in Table 10. For ease of comparison, all contagion shifters are converted to z-scores, such that they have means of 0 and standard deviations of 1. In our preferred final specification, the average level of contagion Γ_{ijt} is 0.4, but with material heterogeneity within that.

The equity of bank i significantly moderates the impact of its exposures on risk. This introduces wide heterogeneity in contagion across banks: a one standard deviation increase in equity reduces contagion by 0.28 in our preferred specification, equivalent to 70% of mean contagion. This is reassuring, and intuitive: all else equal banks with greater loss absorbing capacity should be less sensitive to the risk of their counterparties.

Spillovers are lower between pairs of firms with more closely correlated fundamentals. Once again, the impact is large: a 1 standard deviation increase in our measure of similarity in banks' fundamentals reduces contagion by 0.5 in the richest specification. As explained in Section 4.1.1, dissimilar banks could enjoy risk sharing benefits, but links between more similar banks could also be safer because of better recovery rates or risk identification and management. Our results suggest that the benefits of similarity for contagion exceed the costs. That is, exposures between dissimilar banks are riskier, with any benefits of risk sharing outweighed by potentially lower recovery rates.

We find that contagion is greater when derivatives make up a larger share of an exposure, though this is of borderline significance. This suggests that all else equal derivatives exposures tend to be riskier than exposures via securities financing transactions or fixed income. As discussed in Section 4.1.1, the relative complexity of derivatives or their longer maturities could drive this result. Finally, VIX – a times series proxy of market volatility – has no statistically significant impact on contagion.

In combination, these results suggest that network spillovers are heterogeneous in a way that is material and intuitive given variation across banks and pairs of banks in these contagion shifters. We find particularly strong effects for equity and similarity, which both reduce contagion.

This is important for three reasons. First, it sheds light on how contagion between banks works in practice: we find systematic evidence that some interbank exposures matter more for risk spillovers than others. Second, it sheds light on how banks form networks in response to this: there is evidence of homophily in network formation (bigger links between banks that face similar risks, Elliott et al. (2021)), which we show comes partly from the fact such links are actually less risky. Third, in our results below we show that this heterogeneity matters for identifying systemically important banks and for understanding how banks respond to caps.

5.1.1 Robustness

Here we consider the robustness of our key specification above to the following variations in approach.

First, we add ij fixed effects to our existing it fixed effects, so as to control for time-invariant factors driving exposures between bank pairs. Our chosen contagion shifters all vary at the it or ij level, meaning this combination of fixed effects ensures they do not introduce bias.

Second, we add both jt and ij fixed effects to our existing it fixed effects. This controls for time-varying factors that drive demand by the j bank, and effectively saturates the set of fixed effects available in network formation. Our identification is driven solely by the co-movement of fundamentals and exposures within this demanding set of fixed effects. Suppose, for example, that there were no contagion between banks, but that banks that were doing badly simply demanded additional exposures from others. jt fixed effects control for this type of unobserved variation in demand, giving greater confidence that what we identify is risk contagion.

Third, we remove the exposures scaling for subsidiary banks described in Section 4.1.2, and estimate our parameters based on the raw exposures data. Scaling exposures in the way we do is a reasonable way of dealing with our partial observation of the network, and produces reasonable results, as set out in Table 11. We find, for example, that there is more cross-sectional variation in our estimated a_{it} than inter-temporal variation, which is intuitive. Nevertheless, this robustness test confirms that this exposures scaling is not driving our results.

The results of these tests are reported in Table 12. We find that the main results above continue to hold: greater equity and similarity decrease contagion across all our specifications. Coefficients on exposure characteristics and on VIX are not robustly significant. Our estimates of ϕ are robustly positive and significant, confirming that exposures are costly to banks and that banks respond to increased counterparty risk by shrinking their exposures. The average of contagion across all bank pairs and time also remains robustly positive across all specifications.

5.2 Equilibrium effects

We now turn to the joint determination of risk and exposures in equilibrium. In particular, we use our estimated model to answer the following questions. How does a bank's default risk depend on its exposures? How do its exposures respond to changes in its default risk? What impact does this have on the default risk of other banks? The answer to these questions depends on the parameters estimated and discuss above, and the strength of the equilibrium forces captured by our model. We proceed on the basis of our preferred specification, in the final column of Table 10.

We begin by quantifying the role of contagion in driving bank default risk. Given estimated parameter values for the final period in our sample, we simulate an exogenous 1% increase in all bank exposures and solve for the resulting vector of default risks. The effect of this depends on estimated network spillovers. Given our estimates, we find that mean bank default risk increases by 0.23%. Of this, 0.18% comes from the direct impact of increasing a bank's direct exposures on its risk and the remainder come from the indirect effects that transmit via higher order links in the network. These results confirm the quantitatively large role that interbank exposures play in driving bank default risk.

We now quantify the role of risk in driving exposures. Here we simulate an exogenous 1% increase in all banks' default risks and solve for the resulting matrix of banks' exposures. This effect depends on the magnitude of contagion Γ_{ijt} and the sensitivity of banks' responsiveness to them ϕ. We find that in aggregate banks shrink their exposures by 0.64% in response to this shock. This demonstrates that banks consider risk in determining their interbank exposures, and that this channel is economically meaningful.

Having quantified the marginal impacts of exposures on risk and risk on exposures, we now study them jointly. We do so in two ways. First, we study the equilibrium that would result if banks did not take into account risk when determining their exposures, by setting the final term in equilibrium Equation 5 to zero. The impact on aggregate exposures, surplus and risk is shown in the second row of Table 13. Aggregate exposures increase by 6%, as the moderating impact of risk on exposures is removed. This increase in exposures leads banks' risk to increase by 10%. This is costly for banks, who see their overall surplus fall by 4%.

Second, we study the equilibrium that would result if there was no contagion, by setting $\Gamma_{ijt} = 0$ and re-solving our model. The results of this are shown in the third row of Table 13. Given estimated contagion is large, removing it significantly reduces bank risk by 18%. Banks respond to this reduction in risk by increasing exposures by 31%, and benefitting from higher surplus as a result.

Together, these results highlight that risk and interbank activity are jointly determined, with each playing a large equilibrium role in shaping the other.

5.3 Implications for policy

We have established that contagion is material, that it varies across banks according to their level of loss absorbing capacity and across bank pairs according to the similarity in their fundamentals, and that the network of interbank exposures adjusts in response. We now demonstrate the implications of these facts for policy and efficiency. We first show that these facts imply that looking at exposures alone is insufficient to understand the importance of a bank in a network, with implications for identifying systemically important banks for regulatory purposes. We then study what these facts mean for policies that regulate banks' exposures to each other. To do so, we first demonstrate how a social planner should think about the trade-off between interbank surplus and bank default risk in this setting. Based on this, we show how to design regulatory caps on exposures that best manage this trade-off.

5.3.1 Systemic importance

A recurring issue in the network literature is the identification of "important" nodes. A parallel issue in banking regulation is the identification of "systemically important banks", which in part depends on banks' interconnectedness. Banks designated as systemically important receive greater attention from supervisors and more stringent regulatory requirements (Basel Committee, 2014). In each case, the task is to map from observed outcomes for a set of banks to a ranking of importance. In this section, we argue that our results change the way one should think about this mapping.

We illustrate this using one of the simplest measures of centrality: eigenvector centrality. Applying this centrality measure to the network $\mathbf{C_t}$ gives a ranking of which banks are most systemically important in driving bank default risk. If contagion intensity is homogeneous, $\Gamma_{ij} = \Gamma$, then the level of Γ has no impact on this relative ranking. If, however, contagion intensity is heterogeneous, then accounting for this heterogeneity is important when assessing centrality: a more reasonable measure of centrality would be based on the weighted adjacency matrix $\mathbf{\Gamma} \circ \mathbf{C_t}$. Importantly, the effect of this weighting on the ranking of systemic importance is not random noise, because the equilibrium network depends on this weighting: links C_{ij} where Γ_{ijt} is low are more likely to be large, all other things being equal. As a result, assessing centrality based on the raw, unweighted exposures matrix is likely to systematically overstate the relative centrality of more central banks, and understate the relative centrality of less central banks.

In Figure 3, we show that calculating eigenvector centrality based on unweighted $\mathbf{C_t}$ and weighted $\mathbf{\Gamma} \circ \mathbf{C_t}$ lead to quite different rankings of systemic importance. Bank 16, for example, would be identified as the third most systemically important node based on the unweighted network. Based on the weighted network, however, 8 other banks are more systemically important than Bank 16: in other words, Bank 16's links are large because its links are relatively safe. Bank 12's centrality, on the other hand, is understated when looking solely at the unweighted network: in other words, Bank 12's links are relatively small because its links are comparatively unsafe. We do this for eigenvector centrality, but the same point applies to other centrality measures.

This exercise illustrates the importance of recognizing that banks are heterogeneous in the risks they pose to each other, and that the network they form partly takes this into account. Understanding which banks represent the key nodes in the system requires a framework that can capture and quantify these forces, which our estimated model provides.

5.3.2 Efficiency and policy trade-offs

We now set out a framework by which a social planner can evaluate different outcomes, with a goal of evaluating policy. To do so, we need a framework by which the social planner evaluates outcomes. Regulators, for example, specifically intend to "*preserve the benefits of interconnectedness in financial markets while managing the potentially harmful side effects*" (Yellen, 2013).

The market failure in the model comes from banks taking the default risk of other banks as given when they set their exposures. This produces a familiar result that network exposures are too large (Acemoglu et al., 2015), relative to those that maximize aggregate bank surplus. One complication in our setting is that the social planner may not simply maximize aggregate interbank surplus, but may have preferences over bank risk. This would be the case, for example, if excessive bank risk has externalities for the wider economy. One benefit of having a model of the joint determination of exposures and risk is that we can address this directly.

We do this in Figure 4 by tracing out an optimal possibility frontier that trades off aggregate interbank surplus against mean bank default risk. We draw this locus by numerically solving for the exposures matrix that maximizes interbank surplus conditional on mean bank default risk being equal to some critical value. We then vary this critical value to trace out the frontier. A social planner that cares only about interbank surplus would implement the right-hand point on the frontier in Figure 4, whereas a social planner that cares only about minimizing bank risk would implement the lowest point. We evaluate regulation within this surplus-risk space while remaining agnostic about the social planner's preferences over the two.

5.3.3 Caps on exposures

Finally, we evaluate a key piece of regulation of banking exposures, the system of caps on regulatory exposures under the Basel regulations (Basel Committee, 2014). After the global financial crisis banks were subject to caps on their exposures, whereby no single bilateral exposure can exceed 25% of its capital or 15% for exposures between two "globally systemic institutions". We simulate a generic cap on individual exposures on each bank at κ_i, such that $C_{ijt} \leq \kappa_i$ for all j. We start by setting κ_i equal to i's largest exposure, such that it does not bind, and then gradually tighten the cap by decreasing κ_i. The solid line in Figure 5 show the results of this exercise. As the cap is tightened, both interbank surplus and mean bank default risk shrink. Whether the social planner prefers this outcome to the initial equilibrium will depend on the weight they place on risk vs surplus in social welfare.

We also simulate an alternative counterfactual cap on a bank's aggregate exposures rather than individual exposures, such that $\sum_j C_{ijt} \leq \kappa_i$. We trace out the effect of such a cap in risk-surplus space in the dashed line in Figure 5. As with the bilateral cap, both surplus and default risk shrink. However, the aggregate cap always lies to the south-east of the bilateral cap. That is, for any bilateral cap one can find an appropriately chosen aggregate cap that both increases surplus *and* decreases risk, and thus must be preferred by the social planner regardless of their preferences over surplus and risk. In doing so we remain agnostic in the debate about whether banks should be more tightly regulated, and instead identify a change to the nature of that regulation that must improve outcomes.

This result stems from the two key points we make in this paper: interbank links are heterogeneous, and the interbank network is formed endogenously in a way that takes this into account. Banks respond to caps by adjusting exposures. With a bilateral cap, this can involve leakage, whereby banks simply shift their exposures elsewhere. Aggregate caps can more effectively reduce risk by preventing such leakage. Furthermore, exposures are endogenously related to heterogeneous network spillovers: large exposures may be large precisely because their network spillovers are small. A cap on individual exposures may therefore penalize exposures with low spillover and promote smaller exposures with greater spillover, whereas an aggregate cap does not. Both these forces mean aggregate caps can potentially better manage the risk-surplus trade-off faced by banks and regulators.

6 Conclusion

Contagion from direct exposures between banks has been the subject of much academic and regulatory interest, but relatively little in the way of empirical assessment. We make use of novel data on aggregate exposures between banks to study what these exposures mean for risk, and how banks take risk into account when setting exposures. We document reduced-form facts and structurally estimate a model of contagion and network formation.

There are two broad messages from our paper. The first is that contagion is material and heterogeneous. We quantify how bank default risk responds to changes in the exposures network, and what heterogeneity exists across banks. The second is that banks have an incentive to manage their exposures and their risk: in our model this comes from the fact that becoming riskier increases their funding costs. This means that there is an endogenous relationship between equilibrium exposures and heterogeneous contagion, in that less risky links are more likely to be large. This endogenous relationship matters for the identification of systemically important banks, as ranking bank centrality based only the scale of their exposures is likely to be biased. This also matters for the design of regulation, and in particular the extent to which large exposures specifically, and all exposures more generally, should be constrained.

Tables & Figures

Table 1: Exposures summary statistics

	Mean	Std. Dev	5th pctile	Median	95th pctile
Aggregate					
C_{ijt}	285	499	0.0	92	1,288
$\sum_j C_{ijt}$	4,850	5,293	97.5	3,098	16,556
$100 \times \sum_j C_{ijt}/E_{it}$	9	13	0.2	5	37
C_{ijt} *by year*					
2011	210	499	0.0	5	1,020
2012	248	530	0.0	72	1,041
2013	223	413	0.0	71	874
2014	317	471	0.0	127	1,339
2015	319	539	0.0	115	1,441
2016	323	628	0.0	98	1,414
2017	267	404	0.0	105	1,091
2018	237	368	0.0	83	1,025

Notes: This table summarizes our data on exposures. C_{ijt} denotes the exposure of i to j at time t. E_{it} denotes the equity of i at time t. All figures are in £mn. Figures by year are for the last reporting period in each year.

Table 2: Default risk and fundamentals summary statistics

	Default risk p_{it} (bps)		Fundamentals X_{it} (bps)	
Year	Mean	Std. Dev.	Mean	Std. Dev
2011	279	116	181	88
2012	179	58	101	43
2013	120	37	116	92
2014	95	28	209	77
2015	118	30	164	56
2016	102	35	132	74
2017	74	24	109	45
2018	83	25	188	71

Notes: This table summarizes variation in our measures of default risk and fundamentals. The first two numeric columns plot the mean and standard deviation of banks' CDS premia in the final period of each year in our sample. The final two columns repeat this for our measure of banks' fundamentals, described in Section 2.2.

Table 3: Exposures, risk and fundamentals

	Exposure C_{ijt}			
	OLS		TSLS	
	(1)	(2)	(3)	(4)
Risk of j-bank p_{jt}	0.08		-1.2***	
	(0.18)		(0.39)	
Risk of i-bank p_{it}		0.58**		-0.69**
		(0.28)		(0.35)
R^2	0.38910	0.14791	0.37544	0.14627
Observations	6,426	6,426	6,426	6,426
it fixed effects	Yes		Yes	
jt fixed effects		Yes		Yes

Notes: This table reports the results of regressing interbank exposures C_{ijt} on the risks of each counterparties. Exposures are in £mn and risk denotes CDS premia in bps. The first two specifications are OLS. The final two columns instrument for risk p_{it} and p_{jt} with their fundamentals X_{it} and X_{jt}, described in Section 2.2. Standard errors are clustered at the *it* and *jt* levels. ***,** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 4: Default risk and fundamentals

	Default risk p_{it}			
	(1)	(2)	(3)	(4)
Fundamentals X_{it}	0.75***	0.32***	0.39***	0.10***
	(0.04)	(0.03)	(0.03)	(0.03)
R^2	0.03054	0.30820	0.71738	0.83221
Observations	378	378	378	378
Bank fixed effects		Yes		Yes
Time fixed effects			Yes	Yes

Notes: This table reports the results of regressing default risk on fundamentals and various fixed effects. All variables are in bps. Standard errors are clustered at the bank level. ***,** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 5: Exposure characteristics

	Mean	Standard deviation Aggregate	Cross section	Time series
Instrument				
Derivatives (%)	61.5	42.1	41.8	18.3
Securities financing (%)	25.0	37.7	37.5	19.6
Fixed income (%)	13.5	29.1	28.5	14.9
Maturity				
Short term (%)	67.1	40.3	39.9	26.9
Medium term (%)	12.5	24.3	23.9	16.2
Long term (%)	20.4	34.8	34.3	22.0

Notes: This table summarizes the characteristics of interbank exposures. The first three rows break exposures down by the underlying instrument type. For each bank pair in each time period we compute their derivatives exposures as a share of their total exposures. We then compute the mean and standard deviation of this across all bank pairs and time periods. In the penultimate column, we compute the standard deviation of this statistic across exposures within a period, and then take the average. In the final column we compute the standard deviation of this statistic across periods for each bank pair, and then average. The second and third row repeat this for securities financing exposures and fixed income exposures respectively. The final three rows repeat this for the maturity of banks' securities financing and fixed income activity. Short-term is defined as less than three month residual maturity, and long-term has residual maturity of over one year.

Table 6: Uniqueness simulations

	Mean	Maximum
Initial discrepancy	0.22	0.33
Difference in solution	9.86e-06	9.89e-06

Notes: This table shows the results of our uniqueness simulations, described in Section 3. Initial discrepancy measures the maximum distance between the exposures from which we start our algorithm and the exposures in the known equilibrium. Difference in solution denotes the maximum difference in the exposures to which the solver converges, relative to our known equilibrium. We report the mean and the maximum of these statistics across our 100 simulations.

Table 7: Estimation inputs: summary stats

		Standard deviation		
	Mean	Aggregate	Cross section	Time series
Base variables				
Exposure C_{ijt} (norm.)	0.04	0.06	0.06	0.02
Risk p_{it} (norm.)	0.19	0.11	0.06	0.09
Fundamentals X_{it} (norm.)	0.30	0.16	0.13	0.12
Capital regulation λ_{it} (%)	5.11	1.46	0.14	1.48
Contagion shifters g_{ijt}				
g^1_{it} =log(Equity, £mn)	10.98	0.75	0.77	0.14
g^2_{ij} =Overlap (%)	72.35	23.00	23.00	
g^3_{ij} =Derivative share (%)	56.82	39.59	39.59	
g^4_{ij} =Securities financing share (%)	23.38	31.11	25.06	
g^5_{t} =VIX index	15.62	3.81		3.81
Base instruments Z_{ijt}				
X_{jt}	0.30	0.16	0.13	0.12
X^2_{jt}	0.12	0.14	0.11	0.08
$X_{it}X_{jt}$	0.10	0.08	0.05	0.06

Notes: This table summarizes all of the inputs into our estimation. The first four rows summarize our main variables: exposures, risk and fundamentals – all normalized such that their maximum value is 1 – and capital regulation. The next section summarizes our contagion shifters. Finally we summarize our base instruments, where we note that the summary statistics for X_{it} and X_{jt} are equivalent. For each variable we summarize the mean and standard deviation across the full panel, together with the average within-period and time-series standard deviation.

Table 8: First stage regressions

	Exposure C_{ijt}	Endogenous term $p_{jt} \sum_m C_{imt}$	Default risk p_{it}
Wald stat: irrelevance	7.1	10.4	14.8
Regression R-squared	0.5	0.9	0.7
FE	it	it	t
Observations	4569	4569	378

Notes: This table summarizes the significance of our instrumental variables in regressions of exposures, the endogenous right-hand side variable in Equation 7, and default risk on their respective instruments. The Wald statistic tests whether all instrumental variables are equal to 0, and the null is rejected at the 1% level in each case. Instruments are detailed in Section 4.3, and relevant summary statistics are provided in Table 7.

Table 9: Homogeneous and heterogeneous contagion

	Hom. contagion	Het. contagion $\tilde{\Gamma}_i$	Het. contagion $\tilde{\Gamma}_j$
Mean contagion Γ_{ijt}	1.032***	0.555***	1.388***
	(0.156)	(0.175)	(0.097)
Std dev. contagion Γ_{ijt}		0.529***	0.841***
		(0.052)	(0.072)
Exposure response ϕ	5.435***	7.001***	10.960***
	(0.428)	(1.338)	(0.857)
Contagion FEs	it	it	it
Default FEs	t	t	t
Contagion Obs.	4569	4569	4569
Default Obs.	378	378	378

Notes: This table reports parameter estimates under three alternative parameterisations of contagion intensity. The first column restricts contagion intensity Γ to be constant across banks and time. The second column allows contagion intensity Γ_i to vary according to the bank that holds the exposures. The third allows contagion intensity Γ_j to vary according to the bank to which the exposure is held. We report the mean contagion intensity and its standard deviation across banks, together with estimates of ϕ. Standard errors are heteroskedasticity robust. ***,** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 10: Parameter estimates

	(1)	(2)	(3)	(4)	(5)
Contagion shifters $\tilde{\Gamma}_q$					
Constant	1.032***	0.600***	0.452***	0.047	0.083
	(0.156)	(0.185)	(0.117)	(0.453)	(0.426)
$\mathrm{Equity_{it}}$		-0.076	-0.090**	-0.261***	-0.253***
		(0.056)	(0.039)	(0.064)	(0.068)
$\mathrm{Overlap_{ij}}$			-0.302***	-0.512***	-0.505***
			(0.055)	(0.135)	(0.145)
Derivatives $\mathrm{share_{ij}}$				0.341*	0.301
				(0.174)	(0.184)
SFT $\mathrm{share_{ij}}$				-0.581	-0.647
				(0.721)	(0.694)
$\mathrm{VIX_t}$					-0.195
					(0.268)
Exposure response ϕ	5.435***	5.137***	4.373***	4.051***	4.127***
	(0.428)	(0.420)	(0.381)	(0.381)	(0.437)
Contagion FEs	it	it	it	it	it
Default FEs	t	t	t	t	t
Contagion Obs.	4569	4569	4569	4569	4569
Default Obs.	378	378	378	378	378

Notes: This table reports parameter estimates for various different parameterizations of contagion. Variables are described in Section 4.1.1 and summarized in Table 7. Shifters are included as z-scores to aid interpretation. Standard errors are heteroskedasticity robust. ***,** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 11: Summary of partial exposure scaling

	Scaling α_{it}
Mean	8.4
Std. Dev	19.6
Std. Dev. (across i)	14.1
Std. Dev. (across t)	8.4
5th pctile	0.5
Median	2.8
95th pctile	32.7

Notes: This table summarizes the scaling variables we compute for the exposures of subsidiaries, α_{it}. The procedure for computing these is set out in Section 4.1.2.

Table 12: Robustness

	Baseline	+ ij FEs	+ ij + jt FEs	No scaling
	(1)	(2)	(3)	(4)
Contagion shifters $\tilde{\Gamma}_q$				
Constant	0.083	0.395*	0.400**	0.000
	(0.426)	(0.231)	(0.192)	(0.166)
$\text{Equity}_{\text{it}}$	-0.253***	-0.268***	-0.268***	-0.166***
	(0.068)	(0.074)	(0.076)	(0.055)
$\text{Overlap}_{\text{ij}}$	-0.505***	-0.464***	-0.491***	-0.434***
	(0.145)	(0.132)	(0.167)	(0.139)
Derivatives share_{ij}	0.301	0.215	-0.528	0.236**
	(0.184)	(0.563)	(0.408)	(0.114)
SFT share_{ij}	-0.647	-0.134	-0.695*	-0.318
	(0.694)	(0.460)	(0.358)	(0.280)
VIX_{t}	-0.195	-0.604	-0.514	-0.006
	(0.268)	(0.477)	(0.371)	(0.133)
Exposure response ϕ	4.127***	4.329***	4.216***	2.875***
	(0.437)	(0.415)	(1.493)	(0.646)
Contagion FEs	it	ij+it	ij+it+jt	it
Default FEs	t	t	t	t
Contagion Obs.	4569	4569	4569	4569
Default Obs.	378	378	378	378

Notes: This table reports the robustness of our key specification in Table 10 to various assumptions. The first column repeats the baseline. The second and third column add more onerous fixed effects. The fourth column removes our scaling of exposures from subsidiaries, described in Section 4.1.2. Variables are defined as in Table 10. Standard errors are heteroskedasticity robust. ***,** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 13: Quantifying the role of contagion

	Mean Default Risk	Agg. Exposures	Agg. Surplus
Baseline	1	1	1
No contagion response	1.10	1.06	0.96
No contagion	0.82	1.31	1.15

Notes: This table summarizes the equilibrium impacts of contagion and banks' responses to contagion on risk, exposures and surplus. In the penultimate row we set the first term on the right-hand side of Equation 5 and set it to zero – implying that banks ignore contagion when determining their exposures – and recompute the equilibrium. In the final row we set contagion to zero, such that exposures have no effect on risk, and thus risk has no effect on the network, and recompute the equilibrium. For each exercise, we show mean default risk, aggregate exposures, and aggregate interbank surplus relative to our baseline estimates.

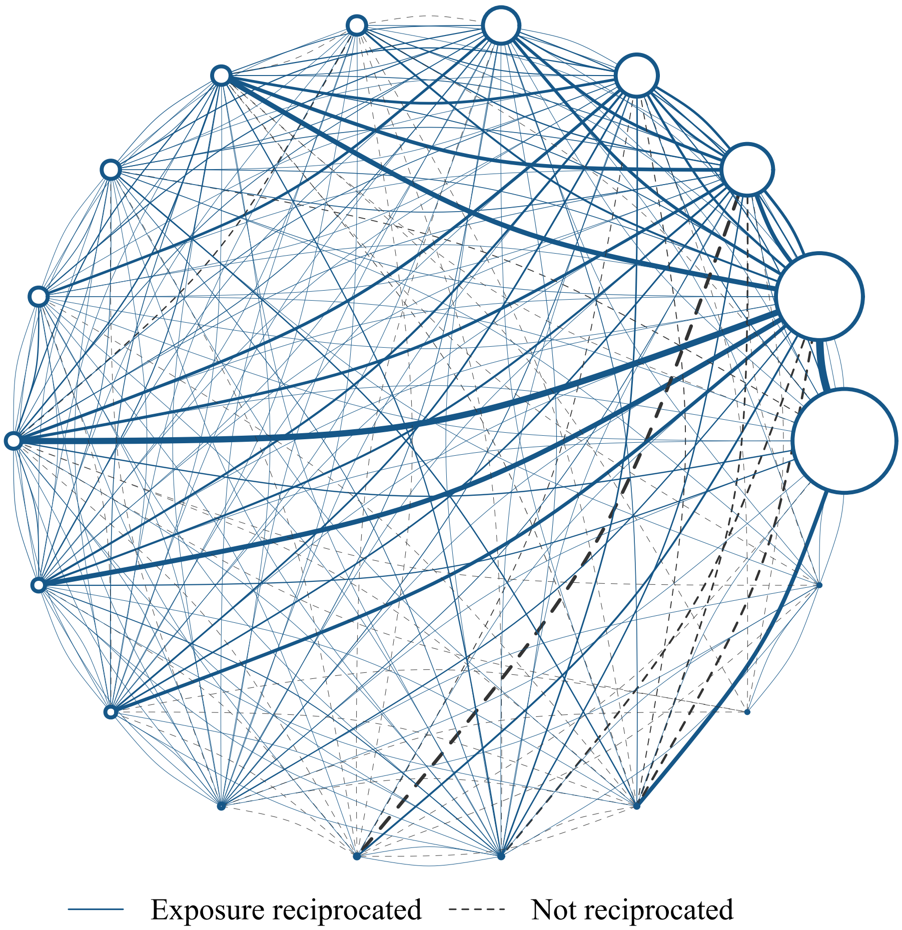

Figure 1: The aggregate network in H1 2015

Notes: This is the network of aggregate exposures between banks in H1 2015. Each node is a bank in our sample. A solid line between two nodes shows a reciprocated exposure (each bank has an exposure to the other) and a dashed lines shows an unreciprocated exposure (that goes in one direction only). The line width is proportional to the size of the exposure. The size of the node is proportional to its total outgoings.

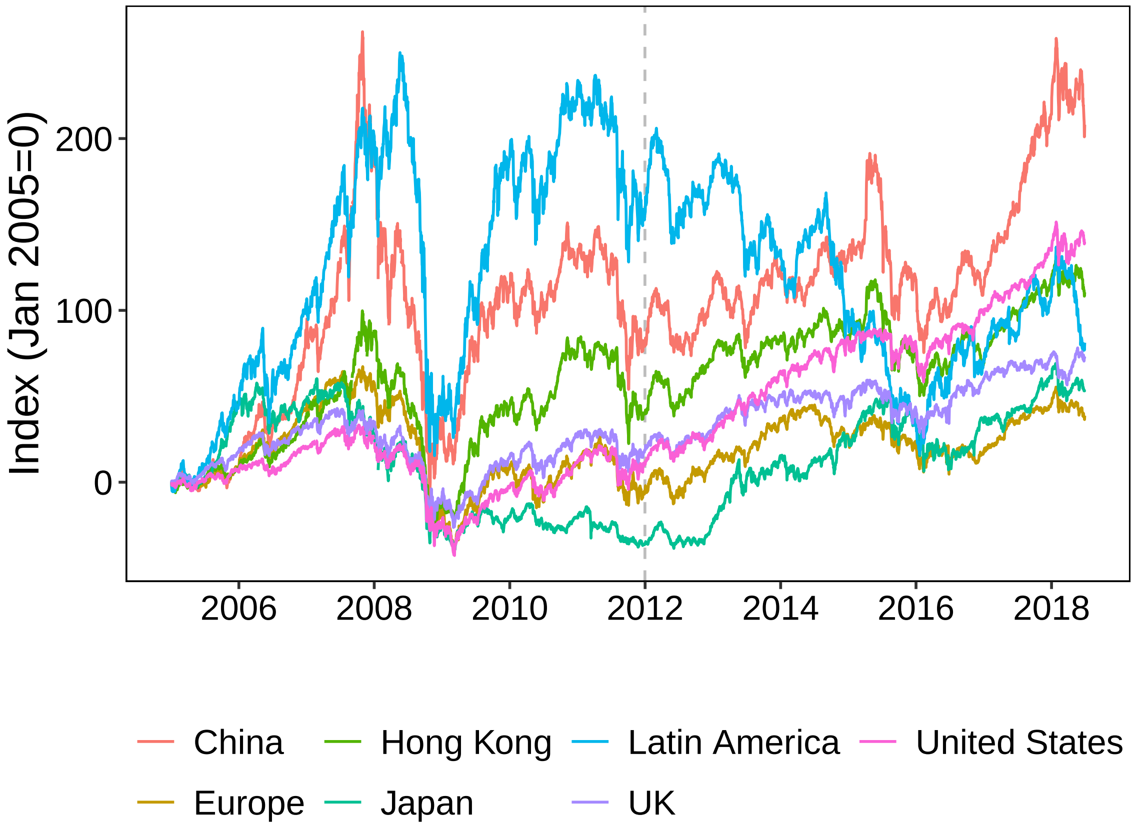

Figure 2: Regional equity indices

Notes: This figure shows variation in our regional equity indices, which we use to compute bank fundamentals. The dashed vertical line indicates the start of our sample period, and the end of our training period when computing fundamentals. Indices are normalized to 0 in 2005.

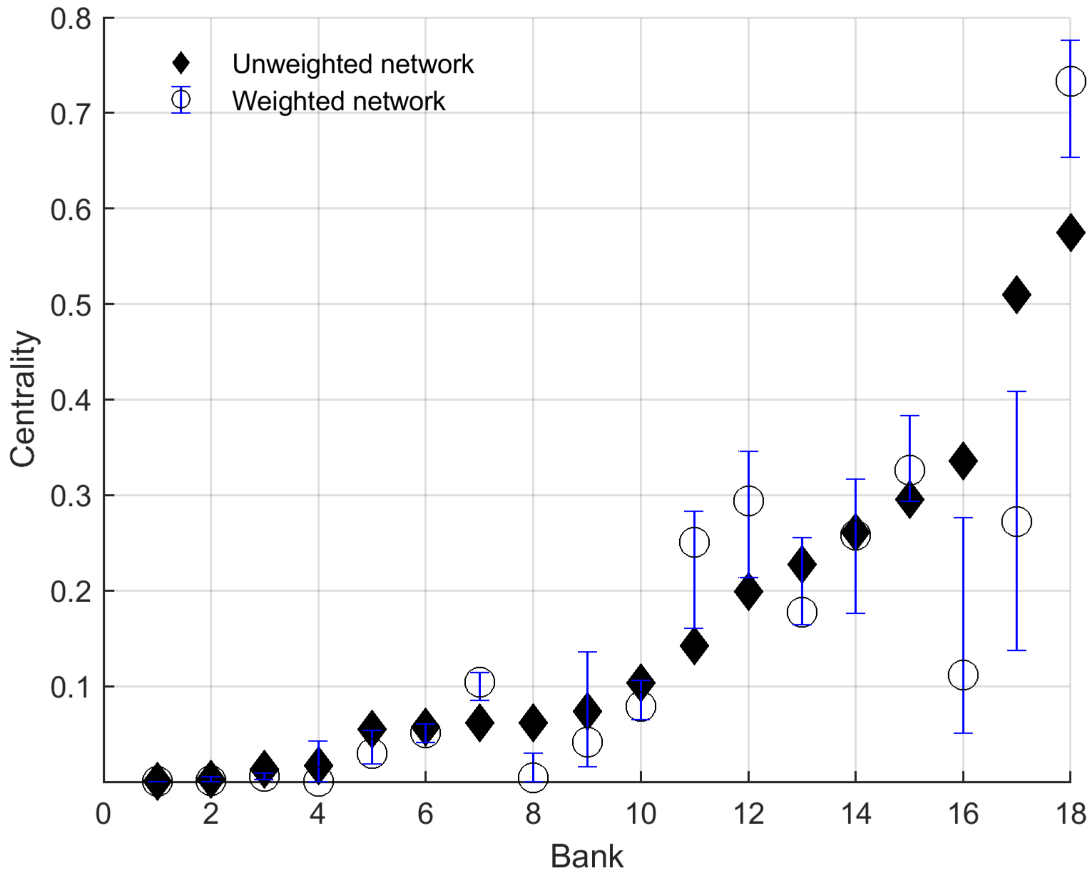

Figure 3: Identifying systemic nodes

Notes: This figure plots the relative centrality of each of the 18 banks in our sample using (right) eigenvector centrality. The black diamonds show relative centrality based on the unweighted network of observed exposures: banks with large exposures are more central. The white circles show relative centrality based on observed exposures weighted by their relative contagion intensities: relatively risky links are given a higher weighting. The blue lines show a 95% confidence interval around this weighted measure.

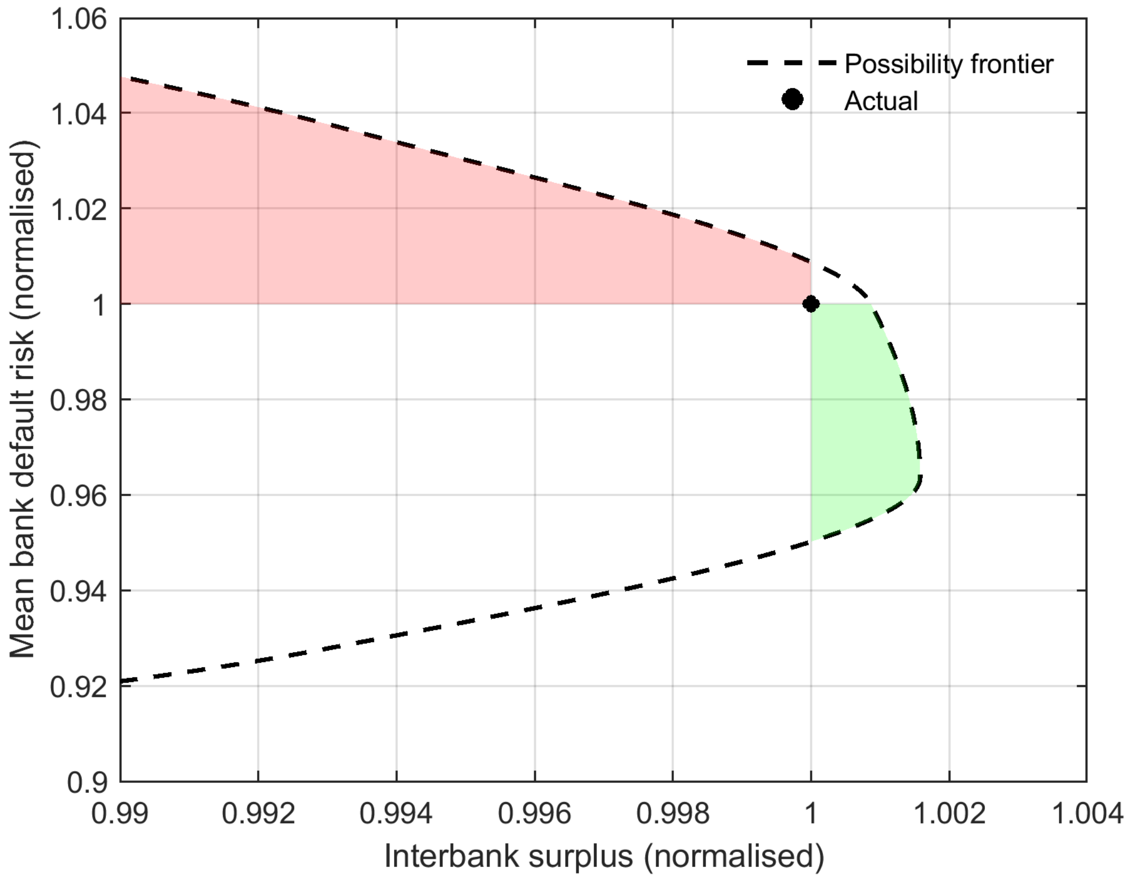

Figure 4: Decentralized efficiency

Notes: This figure quantifies the inefficiency in the decentralized outcome. Point (1,1) shows mean bank default risk and interbank surplus based on actual exposures. The dotted line shows the efficient possibility frontier of combinations of surplus and risk.

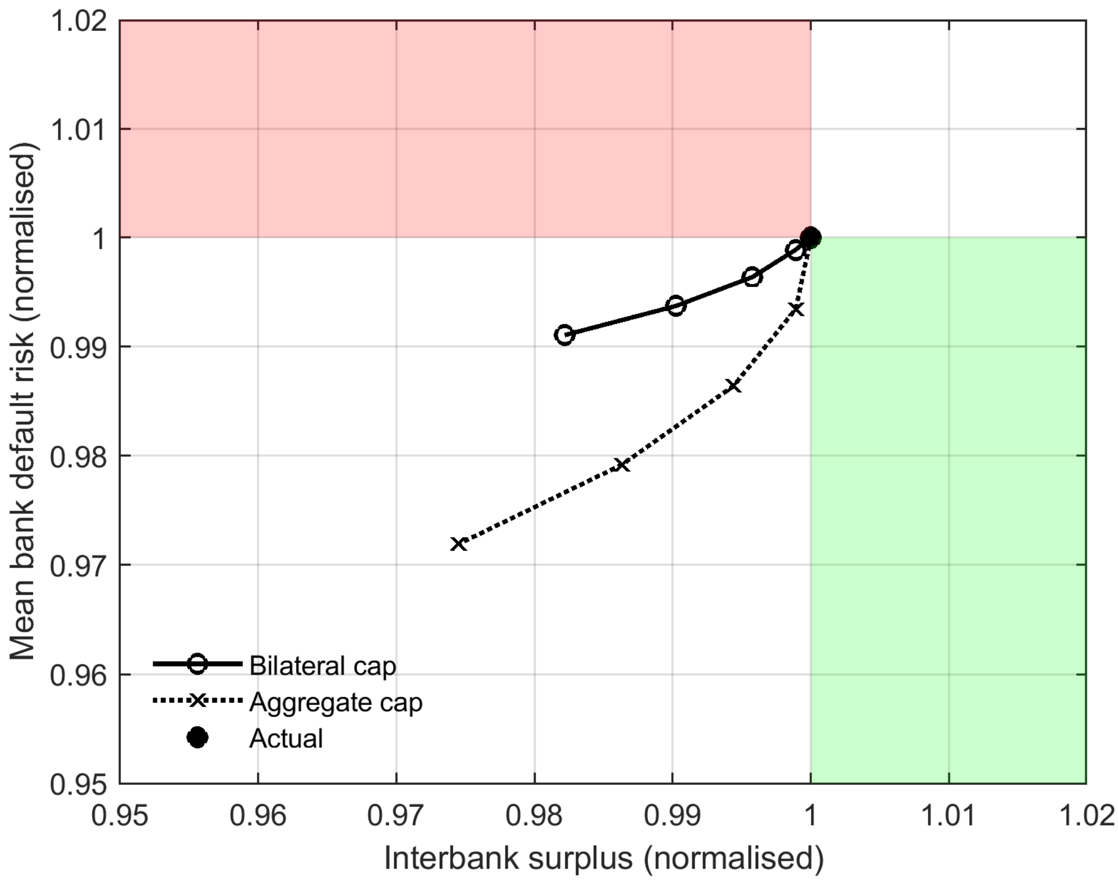

Figure 5: Caps on Exposures

Notes: This figure shows the impacts on default risk and surplus of caps on exposure. We start with actual default risk and interbank surplus (the black circle). We then plot the effect of bilateral caps on banks' exposures, such that no bank can have an exposure greater than $(1-x)\%$ of their existing greatest exposure, for $x = 5\%, 10\%, 15\%$ and 20% (the solid line). We then plot the effect of aggregate caps, such that each bank must have total exposures no greater than $(1-x)\%$ of their existing *total exposures*, for the same increments of x (the dashed line). Aggregate caps can both decrease risk and increase surplus relative to bilateral caps.

References

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Acharya, V. and Bisin, A. (2014). Counterparty risk externality: Centralized versus over-the-counter markets. *Journal of Economic Theory*, 149:153–182.

Acharya, V. V. and Yorulmazer, T. (2008). Information contagion and bank herding. *Journal of Money, Credit and Banking*, 40(1):215–231.

Admati, A. R. (2016). The missed opportunity and challenge of capital regulation. *National Institute Economic Review*, 235:R4–R14.

AFME (2017). CRD 5: The new large exposures framework. Technical report, Association for Financial Markets in Europe.

Afonso, G., Kovner, A., and Schoar, A. (2011). Stressed, not frozen: The federal funds market in the financial crisis. *The Journal of Finance*, 66(4):1109–1139.

Aldasoro, I. and Alves, I. (2018). Multiplex interbank networks and systemic importance: An application to European data. *Journal of Financial Stability*, 35:17–37.

Allen, F., Babus, A., and Carletti, E. (2009). Financial crises: theory and evidence. *Annual Review of Financial Economics*, 1(1):97–116.

Allen, J., Hortaçsu, A., and Kastl, J. (2021). Crisis management in Canada: Analyzing default risk and liquidity demand during financial stress. *American Economic Journal: Microeconomics*, 13(2):243–275.

Allen, J. and Wittwer, M. (2023). Centralizing over-the-counter markets? *Journal of Political Economy*, 131(12):3310–3351.

Antinolfi, G., Carapella, F., Kahn, C., Martin, A., Mills, D. C., and Nosal, E. (2015). Repos, fire sales, and bankruptcy policy. *Review of Economic Dynamics*, 18(1):21–31.

Babus, A. (2016). The formation of financial networks. *The RAND Journal of Economics*, 47(2):239–272.

Bajaj, R., Binmore, A., Dasgupta, R., and Vo, Q.-A. (2018). Banks' internal capital markets: how do banks allocate capital internally? *Bank of England Quarterly Bulletin*, page Q2.

Baker, M. and Wurgler, J. (2015). Do strict capital requirements raise the cost of capital? Bank regulation, capital structure, and the low-risk anomaly. *American Economic Review*, 105(5):315–20.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417.

Basel Committee (2011). Basel III: a global regulatory framework for more resilient banks and banking systems. Technical report, Bank for International Settlements.

Basel Committee (2014). Supervisory framework for measuring and controlling large exposures. Technical report, Bank for International Settlements.

Basel Committee (2018a). Countercyclical capital buffer. Technical report, Bank for International Settlements.

Basel Committee (2018b). The treatment of large exposures in the Basel capital standards. Technical report, Bank for International Settlements.

Batiz-Zuk, E., López-Gallo, F., Martínez-Jaramillo, S., and Solórzano-Margain, J. P. (2016). Calibrating limits for large interbank exposures from a system-wide perspective. *Journal of Financial Stability*, 27:198–216.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the UK mortgage market. *The Journal of Finance*, 76(6):2997–3053.

Benetton, M., Buchak, G., and Garcia, C. R. (2022). Wide or narrow? Competition and scope in financial intermediation. Working paper.

Benetton, M., Gavazza, A., and Surico, P. (2025). Mortgage pricing and monetary policy. *American Economic Review*, 115(3):823–863.

Blasques, F., Bräuning, F., and Van Lelyveld, I. (2018). A dynamic network model of the unsecured interbank lending market. *Journal of Economic Dynamics and Control*, 90:310–342.

Bolton, P. and Oehmke, M. (2015). Should derivatives be privileged in bankruptcy? *The Journal of Finance*, 70(6):2353–2394.

Brancaccio, G. and Kang, K. (2022). Search frictions and product design in the municipal bond market. Working paper.

Cabrales, A., Gottardi, P., and Vega-Redondo, F. (2017). Risk sharing and contagion in networks. *The Review of Financial Studies*, 30(9):3086–3127.

Chang, B. and Zhang, S. (2018). Endogenous market making and network formation.

Coen, J. and Coen, P. (2022). A structural model of liquidity in over-the-counter markets.

Cohen-Cole, E., Patacchini, E., and Zenou, Y. (2010). Systemic risk and network formation in the interbank market.

Craig, B. and Ma, Y. (2022). Intermediation in the interbank lending market. *Journal of Financial Economics*, 145(2):179–207.

De Paula, Á. (2017). Econometrics of network models. In *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, pages 268–323. Cambridge University Press Cambridge.

De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, 12:775–799.

Denbee, E., Julliard, C., Li, Y., and Yuan, K. (2021). Network risk and key players: A structural analysis of interbank liquidity. *Journal of Financial Economics*, 141(3):831–859.

Diamond, W., Jiang, Z., and Ma, Y. (2024). The reserve supply channel of unconventional monetary policy. *Journal of Financial Economics*, 159:103887.

Duffie, D. (2017). Financial regulatory reform after the crisis: An assessment. *Management Science*, 64(10):4835–4857.

Eisfeldt, A. L., Herskovic, B., Rajan, S., and Siriwardane, E. (2023). OTC intermediaries. *The Review of Financial Studies*, 36(2):615–677.

Elliott, M., Georg, C.-P., and Hazell, J. (2021). Systemic risk shifting in financial networks. *Journal of Economic Theory*, 191:105157.

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

Ellul, A. and Kim, D. (2021). Counterparty choice, bank interconnectedness, and systemic risk. Working paper.

European Council (2018). Banking: Council agreement on measures to reduce risk. Technical report, European Council.

Farboodi, M. (2023). Intermediation and voluntary exposure to counterparty risk. *Journal of Political Economy*, 131(12):3267–3309.

Gale, D. and Yorulmazer, T. (2013). Liquidity hoarding. *Theoretical Economics*, 8(2):291–324.

Galeotti, A. and Ghiglino, C. (2021). Cross-ownership and portfolio choice. *Journal of Economic Theory*, 192:105194.

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

Gofman, M. (2011). A network-based analysis of over-the-counter markets. In *AFA 2012 Chicago meetings paper*.

Gofman, M. (2017). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1):113–146.

Goldsmith-Pinkham, P. and Yorulmazer, T. (2010). Liquidity, bank runs, and bailouts: spillover effects during the northern rock episode. *Journal of Financial Services Research*, 37:83–98.

Greene, W. (2018). *Econometric analysis*. Pearson, New York.

Greenwood, R., Stein, J. C., Hanson, S. G., and Sunderam, A. (2017). Strengthening and streamlining bank capital regulation. *Brookings Papers on Economic Activity*, 2017(2):479–565.

Hellmann, T. F., Murdock, K. C., and Stiglitz, J. E. (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 91(1):147–165.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162.

Hull, J. et al. (2009). *Options, futures and other derivatives/John C. Hull.* Upper Saddle River, NJ: Prentice Hall,.

Iyer, R. and Peydro, J.-L. (2011). Interbank contagion at work: Evidence from a natural experiment. *The Review of Financial Studies*, 24(4):1337–1377.

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association*, 66(334):373–374.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

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

Li, Y., Li, Y., and Sun, H. (2022). The network structure of money multiplier. Working paper. (2021-22).

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Nier, E., Yang, J., Yorulmazer, T., and Alentorn, A. (2007). Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060.

Ozdagli, A. and Weber, M. (2017). Monetary policy through production networks: Evidence from the stock market. Working paper.

Pinter, G. and Uslu, S. (2022). Comparing search and intermediation frictions across markets. Working paper.

Shu, C. (2022). Endogenous risk-exposure and systemic instability. Working paper. (17-35).

Welfens, P. J. (2011). The transatlantic banking crisis: lessons, EU reforms and G20 issues. In *Financial Market Integration and Growth*, pages 49–126. Springer.

Yellen, J. (2013). Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications. *Board of Governors of the Federal Reserve System, Washington, DC.*

Appendices: For Online Publication

A1 Data Appendix

In this section we explain the steps that we took in cleaning and processing the raw data to be used in estimation.

Throughout our sample period – 2011 to 2018 – a set of the largest banking entities in the UK were required to report to regulators a granular breakdown of their 20 largest exposures to other banks. After different reporting entities from the same banking group and removing entities that rarely reported any large exposures to other banks in our sample, we obtained a sample of 18 large banks reporting their exposures half yearly from 2011 to 2014, and quarterly thereafter.

We took the following steps in cleaning and processing our data. First, we manually searched through each of our 18 banks' reports for exposures to other reporting banks, and cut the sample down to these interbank exposures. Each bank reports exposures to their top 20 banking counterparties. In 2014 the reporting requirements were changed in two ways: banks were no longer required to report exposures smaller than £10mn, and were required to report their exposures to eight specific UK banks even if they were not in their top 20 bank counterparties. To ensure consistency in our panel, we apply the minimum threshold ourselves to all data before 2014, and after 2014 delete exposures to banks that are not in the top 20 bank counterparties.

For each pair of banks, we compute an aggregate exposure measure across instruments, which aims to account for credit mitigation techniques – such as netting and collateralisation – and various regulatory adjustments intended to represent counterfactual variations in exposures values or collateral in the case of default. For fixed income exposures, we take current mark-to-market exposure. For derivative exposures we take the *exposure at default* and for securities financing transactions we take the *potential exposure*, each of which incorporate regulatory adjustments designed to capture adverse market movements that can increase exposures in case of default.[7]

In some cases we faced missing data, for a specific bank reporting a specific exposure in a specific period. Our approach requires complete information on the exposures, with no missing data, so in these cases we interpolated the data. In particular, in some cases in a

[7] For, see https://www.newyorkfed.org/medialibrary/media/banking/regrept/IDH_Guidelines.pdf.

specific time period, a bank's report for a given exposure is missing or implausible. In these cases we interpolate between the previous and subsequent periods. This affects a very small number of our exposures. In some periods certain banks do not report potential exposure – a measure of after-collateral exposure – for SFTs. In these cases, we scale up the notional – a measure of pre-collateral exposure – by the average ratio of potential exposure to notional that that bank reports in other periods.

A2 Institutional details

In this section we provide further details on the two key banking regulations which we study in this paper: caps on exposures and capital regulation.

In 2014 the Basel Committee on Banking Supervision (BCBS) set out new standards for the regulatory treatment of banks' large exposures (Basel Committee, 2014, 2018b). The new regulation, which came into force in January 2019, introduces a cap on banks' exposures: a bank can have no single bilateral exposure greater than 25% of its capital (where there the precise definition of capital, in this case "Tier 1 capital", is set out in the regulation (Basel Committee, 2014, 2018b)). For exposures held between two "globally systemic institutions", as defined in the regulation, this cap is 15%. These requirements represent a tightening of previous rules, where they existed. For example, in the EU exposures were previously measured relative to a more generous measure of capital and there was no special rule for systemically important banks (AFME, 2017; European Council, 2018).

Banks are subject to capital requirements, which mandate that their equity (where the precise definition of capital, Common Equity Tier 1, is set out in the regulation) exceeds a given proportion of their risk-weighted assets. In 2013 all banks in our sample faced the same capital requirement per risk-weighted unit, λ_i, which was 3.5%. We use the minimum capital requirements as published by Basel Committee (2011) as the minimum requirements for banks. National supervisors can add discretionary buffers on top of these requirements, which we do not include in our empirical work. Since the financial crisis, regulators have changed capital requirements in three ways. First, and most importantly, the common minimum requirement that applies to all banks has increased significantly. Second, capital requirements vary across banks, as systemically important banks face slightly higher capital requirements than non-systemically important banks. Third, capital requirements vary countercylically, in that in times of financial distress they are slightly lower (Basel Committee, 2018a). The result of these changes is that mean capital requirements for the banks in our sample has increased significantly, from 3.5% to over 9% in 2019. There have also been changes to the definition of capital and the measurement of risk-weighted assets, with the general effect of making capital requirements more conservative.