A structural model of interbank network formation and contagion

Patrick Coen†      Jamie Coen‡

January 2019

Abstract

The interbank network, in which banks compete with each other to supply and demand differentiated financial products, fulfills an important function but may also result in risk propagation. We examine this trade-off by estimating a model in which banks form interbank network links endogenously, taking into account the effect of links on default risk. We find that the decentralised interbank market is not efficient: a social planner would be able to increase surplus on the interbank market by 28.8% without increasing mean bank default risk or decrease mean bank default risk by 12.5% without decreasing interbank surplus. We then show that current regulation (caps on exposures and bank-specific capital requirements) is inferior to simple alternative forms of regulation that are more closely targeted at the underlying inefficiencies.

Keywords: Contagion, Systemic Risk, Interbank Network, Network Formation.

JEL classification: L13, L51, G280, G18.

Preliminary and incomplete: Please do not cite, circulate or publish without prior permission from the authors. Any views expressed are solely those of the authors and do not represent those of the Bank of England or state Bank of England policy.

---

*We are particularly grateful to Alessandro Gavazza and Christian Julliard for many helpful discussions. We are also grateful for comments by Aureo de Paula, Glenn Ellison, Angelo Mele, Wolfgang Ridinger, Claudia Robles-Garcia, John Sutton, Kathy Yuan and Ali Yurukoglu, as well as seminar participants at the London School of Economics and the Bank of England. Both authors acknowledge the financial support of the Economic and Social Research Council.

†London School of Economics, p.coen@lse.ac.uk.
‡London School of Economics & Bank of England, w.j.coen@lse.ac.uk.
1 Introduction

Direct interconnections between banks are important in two ways. First, these interconnections fulfill a function, in that there are gains to trade. The interconnection could, for example, involve providing liquidity or acting as the other party in a hedging transaction, which may result in surplus on both sides of the trade. Second, interconnections can open up at least one side of the transaction to counterparty risk: a lender, for example, runs the risk that the borrowing bank will not pay it back. Both effects were important during the crisis and remain important today, and consequently there is significant debate about optimal regulation in this context.\textsuperscript{1}

We consider the following two fundamental economic questions. How do banks form the interbank network,\textsuperscript{2} given the effect of such exposures on their default risk? What inefficiencies exist, including effects within the interbank network and outside it? The answers to these economic questions then lead us to two questions about regulation. Given equilibrium responses by banks, is regulation effective in reducing default risk? If it does reduce default risk, does it do so efficiently in a way that preserves interbank surplus?

We answer these questions by estimating a structural model that includes interbank network formation and network outcome (in this case, the propagation of default risk along the interbank network). In the network formation part of our model, banks compete with each other to supply and demand differentiated financial products on the interbank market. A bank that supplies a financial product to another bank receives a return, but also acquires an exposure that makes it riskier, via the default risk part of the model. We estimate this model based on novel, rich Bank of England data on interbank exposures, and show that the model fits the data well.

We are the first, to our knowledge, to estimate a structural model of the trade-off between default risk and interbank surplus.\textsuperscript{1} Yellen (2013).

\textsuperscript{2}The term the “the interbank market” is typically used to describe short-term (often overnight) lending between banks. Here we use it more generally to cover any form of direct interconnection between banks.
surplus on the interbank market and the causal effect of the interbank market on bank default risk. This allows us to draw novel conclusions about the efficiency of the interbank market and optimal regulation in this context.

We use novel Bank of England data on interbank exposures. These data are collected by the Bank of England through periodic regulatory surveys of banks from 2012 to 2018, in which they report the exposures they have to their major banking counterparties. The data contains rich detail on the specific types of financial product that form the exposure. We argue that it is the most complete data on total exposures in the literature, which allows us to consider questions relating to how exposures affect default risk. The data are incomplete in certain respects regarding geographic coverage and missing banks, with implications for estimation and identification.

The data reveal certain trends including contraction, consolidation and pairwise persistence in exposures over our sample. Most importantly, the data also show significant variation across pairs in the mix of financial products being traded: different banks are supplying and demanding different things. This variation guides our modelling and estimation in that we emphasise heterogeneity across pairs wherever possible.

Our model consists of three parts: the default risk process, demand for interbank financial products and their supply. We model the default risk process as being spatially autocorrelated, but with a generalisation: the effect of exposures on default risk (which we term contagion intensity) is assumed to be heterogeneous across links. That is, some links are more risky than others.

Banks demand interbank financial products to maximise profits from heterogeneous technologies that take these differentiated interbank products as inputs. Banks supplying financial products face a trade-off between the returns they get from supplying financial products and the effect of supplying on their default risk (via the default risk process described above). An increase in a bank’s default risk increases the cost it incurs to access capital markets in
two ways: by increasing its cost of equity and by increasing the price it pays to be supplied financial products on the interbank market. In practice, the model is essentially a multi-product Cournot game with a non-linear cost function with cost externalities across products and across competing firms.

Equilibrium trades and prices depend in an intuitive way on the key parameters of the model: (i) variation in contagion intensity is a key driver of link formation; links form where they are least risky, (ii) risky banks pay more to be supplied financial products because of contagion and (iii) risky banks supply less, as their cost of doing so is higher. Sources of market failure include market power, inefficient allocation of outputs given cost differences and network externalities. We show that our model can replicate the key summary stats in our data, as well as some additional stylised facts from the financial crisis.

The central part of the network formation model considers how equilibrium exposures change as the riskiness of banks changes. We identify this relationship by exploiting variation in regional equity indices as an instrumental variable for bank risk: for example, we take a shock to a Japanese equity index as a shock that affects the Japanese banks and European banks in our sample differently. We argue that this shock is plausibly exogenous when we include the rich set of fixed effects that our panel network data enables (that is, we can control for any unobserved variation that varies across banks but not across pairs): validity requires only that these indices are not correlated with unobserved pairwise variation in the interbank market.

The central part of the default risk process considers how default risk changes as exposures change. In contrast to large parts of the network econometric literature, when estimating this relationship we consider the endogeneity of the network directly. The default risk process is, by assumption, linear in the fundamentals of banks, but our network formation game shows that equilibrium network links are non-linear functions of bank fundamentals. We

---

3See De Paula (2017) for a summary.
therefore use non-linear variation in bank fundamentals as instruments for equilibrium links in the default risk process.

We estimate our model and show that it fits the data well: the $R^2$ for the network formation game is 0.71. We then use our estimated results to answer the key questions set out above. First, we consider the efficiency of the decentralised interbank market, which we do by deriving an efficient possibility frontier that shows the optimal trade-off between interbank surplus and bank default risk. We find that the decentralised interbank market is not on the frontier: a social planner would be able to increase interbank surplus by 28.8% without increasing mean bank default risk or decrease mean bank default risk by 12.5% without decreasing interbank surplus. This result is driven by the fact that our empirical results indicate that (i) banks supplying exposures have market power and (ii) network externalities are significant.

Second, we use our model to simulate proposed regulation, including a cap on individual exposures and an increase in regulatory capital requirements. We find that a cap reduces interbank surplus without affecting mean bank default risk, as banks simply shift their supply to uncapped links. We find that an increase in capital requirements decreases mean bank default risk, but at the cost of reduced interbank surplus.

We set out two alternative forms of regulation that, to our knowledge, are not currently under consideration. The first is a cap on aggregate supply for each bank, rather than a cap on individual exposures. We find that such a cap does reduce mean bank default risk, as it prevents banks moving capped supply elsewhere in the network. We also consider a pairwise adjustment to capital requirements based on their heterogeneous effects on contagion: links that are particularly risky have greater greater capital requirements, and links that are not risky have lower capital requirements. This targets regulatory intervention more closely at the network externalities that are one of the drivers of inefficiency in our model, and so strictly dominates the general increase in capital requirements currently being considered.
We discuss related literature below. In Section 2, we introduce the institutional setting and describe our data. In Section 3, we set out model. In Section 4, we describe our identification strategy. In Section 5, we set out our results. In Section 6, we undertake counterfactual analyses.

1.1 Related literature

Our work is related to three strands of literature: (i) endogenous network formation in financial markets, (ii) the effects of network structure on outcomes in financial markets and (iii) optimal regulation in financial markets.

There is a large theoretical literature on network formation in financial markets, but little empirical work. Our contribution is that we are the first, to our knowledge, to structurally estimate a model of network formation in which banks trade off gains to interbank trade against contagion. Importantly, this allows us to quantify the extent of inefficiency in the market, in terms of both (i) under-exploited gains to trade and (ii) the trade-off between gains to trade and contagion.

There is an extensive literature on the effect of network structure on outcomes in financial markets, both theoretical and empirical. Our contribution is that the network formation part of our model allows us to more clearly identify network effects and consider more realistic counterfactual scenarios. Regarding identification, a structural network formation model allows us to directly address the endogeneity of the network and consider the impact

4 Babus (2016).
5 Cohen-Cole et al. (2010), Craig and Ma (2018).
6 Acemoglu et al. (2015), Ballester et al. (2006) and Elliott et al. (2014)
of unobserved network links.\textsuperscript{8} Regarding counterfactuals, various papers\textsuperscript{9} adjust the network arbitrarily (usually by simulating a failure) and show the impact on market outcomes holding network structure otherwise fixed. In our model, network structure responds endogenously to such a change. Furthermore, the fact we allow for heterogeneity in contagion intensity means that our identification of systemically important banks in the network is quite different from the standard method.\textsuperscript{10}

There is a specialist literature regarding optimal regulation in financial markets.\textsuperscript{11} Our primary contribution is that by considering bank default risk we are able to evaluate bank regulation comprehensively. Various papers consider the effect of bank regulation on outcomes in specific markets (such as bank lending, repo or bond markets),\textsuperscript{12} but without considering bank default risk (which was arguably the primary focus of much recent banking regulation) it is not possible to draw any conclusions about whether regulation is optimal.

\textsuperscript{8}In this respect, our work has similarities to the network reconstruction literature (De Paula et al. (2018), Diebold and Yılmaz (2014)), although our approach is slightly different. The majority of this literature infers a fixed (although sometimes estimated on a rolling basis), unobserved network from observed outcomes, whereas we identify the parameters of a network formation game from the observable part of the network and outcomes, and use those parameters to infer the unobserved part of the network. Our approach has two advantages: it grounds network reconstruction in theory, and it benefits from the additional identifying information contained in the observed part of the network.

\textsuperscript{9}Eisfeldt et al. (2018) and Gofman (2017), for example.

\textsuperscript{10}Systemic importance, in this context, relates to the contribution of an individual bank to aggregate default risk of banks. Various measures of systemic importance exist that depend on the topology of the network. Our model shows how relying solely on the presence and size of links within a network can incorrectly characterize a bank’s importance because of heterogeneous contagion intensity: a link is large (and therefore ostensibly the banks involved are systemically important) because it has low contagion intensity. In other words, bigger links are safer on a per-unit basis. We propose instead measuring systemic importance based on the observed network weighted by contagion intensity, which gives quite different rankings among the banks.

\textsuperscript{11}Adrian and Brunnermeier (2011) and Brunnermeier (2009).

\textsuperscript{12}Including Kashyap et al. (2010) on bank lending, Kotidis and Van Horen (2018) on the repo market and Bessembinder et al. (2018) and Adrian et al. (2017) on bonds.
2 Institutional setting and data

2.1 Institutional setting

The aftermath of the 2008 financial crisis saw a number of regulatory changes aimed at making banks more resilient. Most relevant to our purposes are the changes to the treatment of large exposures, and the increases in capital requirements.

2.1.1 Large exposures cap

In 2014 the Basel Committee on Banking Supervision set out new standards for the regulatory treatment of banks’ large exposures (BCBS, 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.\textsuperscript{13} For exposures held between two “globally systemic institutions”\textsuperscript{14} 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).

2.1.2 Capital requirements

Banks are subject to capital requirements, which mandate that their equity\textsuperscript{15} exceeds a given proportion of their risk-weighted assets. Additional equity in principle makes the bank more robust to a reduction in the value of its assets. The total amount of capital that a bank is required to hold to cover a given risk-weighted asset is the product of the value of the asset, its risk-weighting (which we denote $\rho$) and the capital requirement per unit of risk-weighted

\textsuperscript{13}Where the precise definition of capital, “Tier 1 capital”, is set out in the regulation.

\textsuperscript{14}As defined by the Committee.

\textsuperscript{15}Where the precise definition of capital, Common Equity Tier 1, is set out in the regulation.
asset (which we denote $\lambda$).

The risk-weights, $\rho$, can be calculated using banks’ internal models or based on a standardised approach set out by regulators. The standardised approach is based on the credit rating relevant to the asset: for the significant majority of interbank transactions this will be AAA, the highest credit rating.

In 2013, all banks in our sample faced the same capital requirement per risk-weighted unit, $\lambda$, which was 3.5%. Since then, regulators have implemented some minor variation between banks ($\lambda$ is marginally higher for systemically important banks) and some state-specific variation depending on the global economy (a countercyclical capital buffer means that $\lambda$ is lower in times of financial distress (BCBS, 2018a)). Regulators have also implemented a general increase in $\lambda$ across all banks: mean $\lambda$ in our sample is set to increase from 3.5% in 2013 to 10.5% by 2019, as set out in Figure 5.

2.2 Data

2.2.1 Exposures

We use regulatory data on bilateral interbank exposures, collected by the Bank of England. 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 debt instruments, securities financing transactions and derivative contracts. The data include granular breakdowns of each of their exposures: by type (e.g. they break down derivatives into interest rate derivatives, credit derivatives etc.), currency, maturity and, where relevant, collateral type.

We use this dataset to construct a series of snapshots of the interbank market 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 in debt instruments, securities financing transactions,
and derivatives. The result is a panel of $N = 18$ banks over $T = 23$ periods from 2011 to 2018 Q2, resulting in 7,038 ($N(N - 1)T$) observations.

For each $C_{ijt}$, we use the granular breakdowns to calculate underlying “exposure characteristics” that summarise the type of financial instrument that make up the total exposure. These characteristics, which we denote $d_{ijt}$, relate to exposure type, currency, maturity and collateral type. One characteristic of $C_{ijt}$, for example, is the proportion of the total exposure that comes from derivatives.

The dataset offers a unique combination of breadth and detail in measuring exposures. Much of the existing literature (such as Denbee et al. (2017)) 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.

The dataset does have limitations. First, whilst the dataset includes most of the world’s largest banks, it omits banks that do not have a subsidiary in the UK\textsuperscript{16} are omitted. Second, 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. Third, the data are censored: we only see each bank’s top 20 exposures, and only if they exceed £5 million. Finally, the data does not include prices or interest rates.

2.2.2 Default risk

We follow Hull et al. (2009) and Allen et al. (2011) in calculating the (risk-neutral) probability of bank default implied by the spreads on publicly traded credit default swaps.

\textsuperscript{16}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.
2.2.3 Other data

We supplement our core data with the following:

- Geographic source of revenues for each bank (annual, Bloomberg). Bloomberg summarises information from banks’ financial statements about the proportion of their revenues that come from particular geographies, typically by continent, but in some cases by country.

- Macro-economic variables from the World Bank Global Economic Monitor, a panel of 348 macro series from a range of countries.

- Commodity prices from the World Bank “Pink Sheet”, which is a panel of 74 commodity prices.

- S&P regional equity indices for US, Canada, UK, Europe, Japan, Asia, Latin America.

2.3 Summary statistics and stylised facts

The data reveal certain stylised facts about how exposures have varied over our sample.

- There has been relatively steady contraction in the exposures market, as set out in Figure 5. This contraction occurred at the same time as increases in $\lambda$.

- There is positive co-movement between median exposures and median default risk, as set out in Figure 6.

- There has been consolidation in the interbank market, in that the proportion of exposures in our sample that are equal to zero has increased, as has the HHI index over exposure supply, as set out in Figure 7.

- There is significant persistence in exposures, as set out in Table 1, in which we show that the $R^2$ for a regression of $C_{ijt}$ on pairwise ij fixed effects and time fixed effects is
0.51. A regression on it or jt fixed effects results in an $R^2$ of 0.18 and 0.14, respectively, indicating that there is significant pair-wise variation in any given time period.

- Finally, there is significant variation in product characteristics, as set out in Table 1: regressing our measures of product characteristics on it or jt fixed effects results in an $R^2$ of under 30%.

Our sample starts in 2012, and so does not feature the financial crisis. We note three features that were observed on the interbank market during the crisis, on the basis that a model that is useful for the evaluation of regulation and stress-testing should be able to replicate what happened during the crisis. First, risky banks were not supplied; in other words, they experienced lockout (Welfens, 2011). Second, risky banks did not supply, which we loosely term liquidity hoarding (Gale and Yorulmazer, 2013). Third, in the worst periods of the financial crisis there was effectively market shutdown, in that very few banks were supplied anything on the interbank market (Allen et al., 2009).

3 Model

3.1 Overview

There are $N$ banks. At time $t$, the interbank market consists of an $N \times N$ directed adjacency matrix of total exposures, $C_t$. $C_{ijt}$ is the element in row $i$ and column $j$ of $C_t$, and indicates the total exposure of bank $i$ to bank $j$ at time $t$. $C_t$ 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 $i$. For each $C_{ijt}$, $d_{ijt}$ is an $L \times 1$ vector of product characteristics.

$p_t$ is an $N \times 1$ vector of bank default risks: the element in position $i$ is the probability of default of bank $i$. $p_t$ is a function of $C_t$ and an $N \times K$ matrix of bank fundamentals, which we denote $X_t$. This function is the default risk process, and the effect of $C_t$ on
\( p_t \) represents “contagion”, as we will define more formally below. Fundamentals update exogenously according to a random walk: \( X_t = X_{t-1} + e_t^X \), where \( e_t^X \) are drawn from some distribution with mean zero.

\( C_{ijt} \) results in profits to bank \( i \) (we term this supply of exposures) and to bank \( j \) (demand for exposures). Bank \( j \)’s profits on the demand-side are a function of \( C_t \) and \( X_t \). Bank \( i \)’s profits on the supply-side are a function of \( C_t \) and \( p_t \).

The timing of the game is as follows:

1. Given fundamentals \( X_{t-1} \), banks trade to form the interbank network \( C_t \).
2. Fundamentals \( X_t \) update exogenously.
3. \( X_t \) and \( C_t \) imply \( p_t \), via the default risk process.
4. \( p_t \) and \( C_t \) determine the payoffs on the demand-side and the supply-side.

We introduce this friction in timing to ensure that contagion occurs in equilibrium.\(^{17}\) When we consider the effect of a change in fundamentals or network structure, we define the short run effect as the effect before banks are able to reform the network and the long run effect as the effect after the network re-equilibrates.

### 3.2 Default risk process

A bank’s default risk process is the sum of two components: a set of fundamentals and a spatially autocorrelated component whereby bank \( i \)’s default risk depends on its exposure to

\(^{17}\)Consider if instead network formation was frictionless, such that once fundamentals update banks are able to change their exposures immediately. In this case, bank \( i \) would immediately reduce \( C_{ijt} \) if bank \( j \) suffered a shock to its fundamentals, such that in equilibrium bank \( i \) would never be harmed by a fall in the fundamentals of its counterparties. We use this particular timing assumption (as opposed to assuming, for example, that bank \( i \)’s choice of \( C_{ijt} \) today affects the cost of capital it incurs on \( C_{ijt+1} \) tomorrow) because it means the problem is static.
bank $j$, $C_{ijt}$, and bank $j$’s default risk, $p_{jt}$. In matrix form:

$$p_t = X_t\beta + \tau_t(\Gamma \circ C_t)p_t + e_t^p$$

where $\beta$ is a $K \times 1$ vector that represents each bank’s loadings on fundamentals $X$, $\Gamma$ is a matrix of contagion intensities $\Gamma_{ij}$, $\tau_t$ is a scalar that allows for contagion intensities to vary across time and $\circ$ signifies the Hadamard product.

This is a standard spatially autocorrelated regression, as is commonly used in network econometrics,\(^\text{18}\) with a generalisation: the intensity of contagion, $\Gamma_{ij}$, is allowed to be heterogeneous across bank pairs. We define contagion from bank $j$ to bank $i$ as $\frac{\partial p_{it}}{\partial p_{jt}} > 0$: that is, the default risk of bank $j$ has a causal impact on the default risk of bank $i$. In our model, $\frac{\partial p_{it}}{\partial p_{jt}} = \tau_t\Gamma_{ij}C_{ijt}$, such that the strength of contagion depends on the size of the exposure.

This is intended to be a simple reduced form\(^\text{19}\) representation of counterparty risk.

$\Gamma$ can be thought of as contagion intensity in that $\Gamma_{ik} > \Gamma_{im}$ 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. This heterogeneity could come from two sources. First, it could be a result of risk sharing, whereby if bank $i$ and $k$ ($m$) have fundamentals that are positively (negatively) correlated then exposure $C_{ik}$ ($C_{im}$) is particularly harmful (benign).\(^\text{20}\) Second, it could be a result of variations in product, $C_{ij}$. This value $C_{ij}$ is total exposure of $i$ to $j$ across various product/instrument types, some of which may be more sensitive to the default risk of $j$: a simple example is senior and junior debt tranches, whereby clearly the latter is more sensitive to the default risk of the issuer. This difference across products could be modelled using a richer default risk process that separately includes exposures matrices for each instrument type with differing contagion

\(^{18}\)De Paula (2017).

\(^{19}\)As opposed to structural or balance-sheet-based models of contagion, such as Brunnermeier (2009).

\(^{20}\)This implies a relationship between the fundamentals processes and $\Gamma_{ij}$ which we leave open for now, but consider in our empirical analysis.
intensities, but this would introduce an infeasible number of parameters to take to data.

If $\Gamma_{ij}$ is the same for any bank pair then this reduces down to a standard scalar spatially autocorrelated model as is common in the network literature (eg, De Paula (2017)).

We allow for contagion intensity to vary across time via $\tau_t$ because there are, in principle, things that could affect contagion intensity. One of the purposes of the increase in $\lambda$, for example, was to make holding a given exposure $C_{ijt}$ safer (because with greater $\lambda$ bank $i$ has a greater equity buffer if bank $j$ defaults). We do not make any assumptions about the relationship between $\tau_t$ and $\lambda$ at this stage, but consider it in estimation.

Subject to standard regularity conditions on $\Gamma$ and $C$ this spatially autocorrelated process can be inverted and expanded as a Neumann series as follows, which we term the Default Risk Process (“DRP”):

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

Our focus is on counterparty risk, and so $p_{it}$ is not directly affected by $C_{jit}$, but only by $C_{ijt}$. In other words, there is no liquidity risk whereby the failure to be supplied exposures causes bank $i$ to be riskier.

### 3.3 Demand

Each $j$-bank has a technology that maps inputs into gross profit, from which the cost of inputs is subtracted to get net profits. Inputs are funding received from other banks $C_{ij}, \forall i \neq j$ and an outside option $C_{0j}$ designed to capture funding from banks outside my sample and
non-bank sources:

\[
\mathbb{E}[\Pi^D_{jt}] = (A_1 + A_2\mathbb{E}[X^b_{jt}]) \sum_{i=0}^{N} C_{ijt} + \sum_{i=0}^{N} C_{ijt}(\zeta_{ij} + e^D_{ijt}) \\
-\frac{1}{2} \left( B \sum_{i=0}^{N} C^2_{ijt} + 2 \sum_{i=0}^{N} \sum_{k \neq i} \theta_{ik} C_{ijt} C_{kjt} \right) \\
- \sum_{i=0}^{N} r_{ijt} C_{ijt}
\]

Bank j chooses \(C^D_{ijt}\) to maximise net profit taking interest rates as given, resulting in optimal \(C^*_{ijt}^D\) such that inverse demand is as follows:

\[
r_{ijt} = A_1 + A_2 X^e_{jt} + \zeta_{ij} + e^D_{ijt} - BC^*_{ijt}^D - \sum_{k \neq i} \theta_{ik} C^*_{kjt}^D
\]

where \(X^b_{jt} \equiv X_j \beta\) is the expected aggregate fundamental of bank j in its default risk process, \(\zeta_{ij} + e^D_{ijt}\) is heterogeneity in the sensitivity of the j-bank’s technology to product i (separated into two parts for convenience when considering identification below), \(B\) governs diminishing returns to scale and \(\theta_{ik}\) governs the substitutability of product i and k.

We assume that the j-bank has an increasing but concave objective function in the funding that it receives. We justify its concavity on the basis that the j-bank undertakes its most profitable projects first (or conversely, if its funding is restricted for whatever reason, it terminates its least profitable projects rather than its most profitable projects). Concavity also means that the returns to receiving funding decrease, in that the j-bank only has a limited number of opportunities for which it needs funding.

The returns that the j-bank gets from funding are time-varying, in that we allow them to depend on the j-bank’s fundamentals, via \(A_2\). \(A_2 > 0\) means that when the j-bank’s fundamentals are bad the payoff to receiving funding is greater, in that the projects being funded are more important (if, for example, it needs this funding to undertake non-discretionary,
essential projects or to meet margin calls on other funding). This is intended to allow for the importance of the interbank market in times of distress. $A_2 < 0$ means that when the j-bank’s fundamentals are good the payoff to receiving funding is greater; because, for example, there are better, more profitable projects to be undertaken when the bank is doing well.

The technologies possessed by each j-bank vary by $\zeta_{ij} + e_{ijt}$, which governs the importance of product of the i-bank to the j-bank’s technology. We allow this technology to be heterogeneous across pairs.

We also allow for product differentiation, in that the product supplied by bank i may not be a perfect substitute for the product supplied by bank k. We parameterise this product differentiation in parameters we denote $\theta_{ik}$. General $\theta_{ik}$ cannot be reasonably estimated from our dataset; instead we parameterise it as being a logistic function of certain product characteristics, including maturity, currency and instrument-type.

$$\theta_{ik} = \frac{e^{\theta_{ij}}}{1 + e^{\theta_{ij}}}$$

where $d_{i,t}$ denotes the value for characteristic l of bank i, $e_{ik}$ denotes random, unobserved variation in product differentiation and $\tilde{\theta}_l > 0$. For instrument type, for example, $d_{i,t=type}$ is the proportion of i’s product that is derivatives. If banks i and k have very different product characteristics, then $\theta_{ik}$ is small and the two are not close substitutes. If, on the other hand, banks i and k have very similar product characteristics then $\theta_{ik}$ is large and the two are close substitutes. This replaces $\theta_{ik}$ (which across all pairs has dimension $N^2$) with $\tilde{\theta}_l$ (which has dimension $L + 1$).

We assume that product differentiation between interbank exposures and the outside demand option is not related to product characteristics (as every type of product characteristic...
should be available in the wider financial market) but rather the fact that the outside option does not come from banks. We assume that the effect of this difference is constant across all banks: \( \theta_{0j} = \theta_0 \).

### 3.4 Supply

Bank \( i \) has an endowment \( E_{it} \) that it can either supply to another bank or to an outside option. When it supplies its product to bank \( j \) it receives return \( r_{ijt} \) and incurs marginal cost \( mc_{ijt} \). We assume that this marginal cost is the cost of the equity that regulatory capital requirements mandate, which we assume is a function of the bank’s default risk:

\[
m_{c_{ijt}} = \lambda_{ijt} \phi_1 p_{it}.
\]

i-bank’s problem in period \( t \) is to choose \( \{C_{ijt}\}_j \) to maximise:

\[
E[\Pi_{it}] = \sum_j C_{ijt} [r_{ijt} - E[mc_{ijt}(p_{it})] + e^S_{ijt}] + \Pi^{D}_i(\{C_{kit}\}_k, p_{it}) + (E_{it} - \sum_j C_{ijt})r_{i0t}
\]

such that \( C_{ijt} \geq 0, E_{it} - \sum_j C_{ijt} \geq 0 \) and \( mc_{ijt} = \lambda_{ijt} \phi_1 p_{it} \).

For interior solutions the first order condition implies the following inverse supply equation for \( i \neq j \):

\[
r_{ijt} = -\frac{\partial r_{ijt}}{\partial C_{ijt}} C_{ijt} + m_c e_{ijt} + \frac{\partial p_{it}}{\partial C_{ijt}} \sum_k \frac{\partial m_{c_{ijt}}}{\partial p_{it}} C_{ikt} - \frac{\partial \Pi^D_i}{\partial p_{it}} \frac{\partial p_{it}}{\partial C_{ijt}} + r_{i0t} + e^S_{ijt}
\]

where the \( e \) superscript is a shorthand for expectations.

Bank \( i \), when choosing to supply \( C_{ijt} \), therefore balances the return it gets from supplying against the effect of its supply on its default risk, via the default risk process described above. Being riskier harms bank \( i \) by increasing the price it pays to access capital in two ways. First, it increases the marginal cost bank \( i \) pays when supplying interbank exposures (the third

\[21\text{For ease of exposition we have collapsed the risk-weighting (\( \rho \), using the notation from Section 2) and the capital required per risk-weighted assets (\( \lambda \)) into a single parameter, \( \lambda \).}
term in the equation above). Second, being riskier means that bank i pays higher interest rates when demanding exposures (the fourth term in the equation above).

We are able to consider demand and supply sequentially because a bank’s demand choices do not affect its supply choices.

### 3.5 Outside options

As described above, banks face two outside options. On the demand side, bank j can receive external funding at rate \( r_{0jt} \). On the supply side, bank i can supply funding externally and receive rate \( r_{i0t} \). For simplicity, and given that we do not have data on these outside options, we assume that these rates are exogenous to the interbank market, although we do allow them to vary with exogenous fundamentals \( X_{jt} \): \( r_{0jt} = \phi_D X_{jt} \) and \( r_{i0t} = \phi_S X_{it} \). We also run robustness checks with other functions involving \( X \) and \( \lambda \).

### 3.6 Equilibrium

Before considering equilibrium, we summarise what our model implies for the definition of a bank. In our model, bank i is the following tuple: \((E_{it}, d_{i,l}, \beta_i, \zeta_i, \Gamma_i)\): respectively, an endowment, a set of product characteristics, a set of loadings on fundamentals, a technology and a set of contagion intensities. In other words, although the model is heavily parameterised, it allows for rich heterogeneity among banks.

**Definition 3.1** In this context we define a Nash equilibrium in each period \( t \) as: \( N \times N \) matrix of demand \( C^{oD} \), \( N \times N \) matrix of supply \( C^{oS} \) and \( N \times 1 \) vector of default risks \( p^* \) such that markets clear and every bank chooses its links optimally given the equilibrium actions of other banks.

For interior solutions where \( C_{ijt} > 0 \), market clearing implies that supply equation and demand equation are equal, giving the following Equilibrium Quantity Condition for \( C_{ijt} \)
(which we denote “EQC”):

\[
0 = A_1 + A_2 X_{jt}^b + \zeta_{ij} + e_{ijt}^D - 2BC_{ijt} - \sum_{k \neq i}^N \theta_{ik} C_{kjt} + e_{ijt}^S - \lambda_{ijt} \phi_1 p_{jt}^f
\]

\[
-\lambda_{ijt} \phi_1 p_{jt}^f - \phi_1 \tau_t \Gamma_{ij} p_{jt}^f \sum_{k \neq i}^N C_{ikt} \lambda_{ikt} - r_{i0t} - \phi_1 \tau_t \Gamma_{ij} p_{jt}^f \sum_{k} C_{kit} \Gamma_{ki} \sum_{m} C_{kmt} \lambda_{kmt}
\]

We show our calculations in Appendix A. For corner solutions where \( C_{ijt} = 0 \), in equilibrium this expression is instead weakly less than zero, indicating that the returns to supplying \( C_{ijt} \) are less than the cost of doing so.

Equilibrium \( p_{it} \) come from DRP, which we repeat for convenience:

\[
p_t = (I - \tau_t \Gamma \circ C_t)^{-1} (X_t \beta + e_t^p) = \sum_{s=0}^{\infty} (\tau_t \Gamma \circ C_t)^s (X_t \beta + e_t^p)
\]

Substituting \( p^* \) out of EQC using DPR gives a system of equations in \( C^* \). The form of DPR is such that the EQC become a system of infinite-length series of polynomials, such that it is difficult to make definitive statements about uniqueness or even existence.

In principle a solution to this system could be found directly using a non-linear solver. We instead propose the following solution algorithm:

1. Define starting point \( p^0 = X \beta \); that is, fundamental default risk excluding contagion.

2. Given this \( p^0 \), solve the EQC for \( C^0 \).

3. Given \( C^0 \), use DRP to update the default risk to include contagion along the network: \( p^1 = (I - \tau_t \Gamma \circ C^0)^{-1} X \beta \).

4. Given \( p^1 \), solve the EQC for \( C^1 \).
5. Iterate updating $p$ and $C$ until convergence.

The only numerical component of this solution algorithm is solving EQC for $C$, given $p$. Conditioning on $p$ in this way means that the EQC is a system of quadratic equations in $C$ rather than a system of higher order polynomials, which has two advantages. First, algorithms exist that solve a system of quadratic equations quickly and accurately. Second, there are clear results about when such a system has a unique solution, such that it is possible to identify when this solution algorithm converges to a unique solution; even if the game itself might have multiple equilibria, this at least means that we can be sure that we are reaching the same equilibrium when running counterfactuals.

### 3.7 Optimal networks

There are three potential sources of inefficiency in our model. First, there are externalities within the interbank network, as bank $k$’s default risk $p_{kt}$ is affected by $C_{ijt}$ provided that bank $k$ has a chain of strictly positive exposures to $i$. If $C_{kit} > 0$ then this is trivially true, but it is also true if bank $k$ has a strictly positive exposure to another bank that has a strictly positive exposure to $i$, and so on. Banks $i$ and $j$ do not fully account for the effect on $p_{kt}$ when they transact, such that this negative externality implies that exposures are too large relative to the social optimum.

Second, the banks supplying financial products may have market power (the extent of which depends on $\theta$ and $N$), such that exposures are too small relative to the social optimum. Third, equilibrium allocations among suppliers may not be efficient, given differing marginal costs.

We also consider impacts outside of the interbank market. The financial crisis indicates that the default risk of banks can impact other agents outside of the interbank market (such as its depositors, creditors, debtors and various other forms of counterparty). In other words, a social planner would not set exposures and default risk solely to maximise surplus in the
interbank market. We do not quantify surplus outside of the interbank market, but only make the following directional assumption:

**Assumption 1** *Surplus outside of the interbank market is decreasing in the mean default risk of banks.*

This assumption allows us to think about optimal default risk and interbank surplus in the sense of Pareto-optimality. That is, denote total surplus in the interbank market by $TS_I$ and mean default probability by $\bar{p}$, and suppose $TS_I^H > TS_I^L$ and $\bar{p}^H > \bar{p}^L$. Assumption 1 implies that $(TS_I^H, \bar{p}^L) \succ^{SP} (TS_I^L, \bar{p}^H)$, where $\succ^{SP}$ denotes the social planner’s preferences, but it does not allow us to rank $(TS_I^H, \bar{p}^H)$ and $(TS_I^L, \bar{p}^L)$, as we illustrate in Figure 1.

It is helpful to think about the trade-off between $TS_I$ and $\bar{p}$ in terms of constrained maximisation of interbank surplus subject to a default risk constraint: $\max_C TS_I(C) \text{ st } \bar{p}(C) < \bar{p}^{CV}$. Varying critical value $\bar{p}^{CV}$ traces out a possibility frontier showing optimal $TS_I$ for any given $\bar{p}$. Figure 1 shows the frontier and illustrates what conclusions we can draw using this model about different outcomes.

Finally, we note that although it is straightforward to consider efficient allocations, it is much more difficult to calculate optimal regulations (in our model, $\lambda^{SP}_{ijt}$, imposed by the social planner) that fully implement efficient allocations. We consider feasible regulations that are efficiency improvements over the perfectly decentralised market in the section below on counterfactual analysis.

### 3.8 Comparative statics

The model is intended to allow as much heterogeneity among banks as possible, which means that comparative statics depend on the parameter values. In Appendix B we show comparative statics using our estimated parameters. Here, we draw some general model implications to argue that the model can match the data in a reasonable way.
We argue that the model can replicate the summary statistics. First, an increase in $\lambda$ may\footnote{An increase in $\lambda$ increases the amount of equity required per unit supplied, but may reduce the cost of that equity per unit if it makes banks significantly less risky through parameter $\tau_t$.} increase the marginal cost of supplying, which would result in the interbank network contracting and consolidating (as in Figures 5 and 7). Second, the unconditional correlation between $C$ and $p$ can be positive (as in Figure 6) either because of contagion or because of time-variation in demand via parameter $A_2$. Third, individual links will be persistent across time (as in Table 1) because of fixed pairwise characteristics ($\Gamma_{ij}$ regarding contagion intensity and $\zeta_{ij}$ regarding technology) and/or persistent fundamentals. Fourth, the fact
that in the majority of pairs in our sample banks supply assets in both directions is difficult to explain with a homogenous asset, but easy to explain with product differentiation.

We also argue that the model can, broadly speaking, replicate the stylised facts from the crisis that we set out above. First, if bank i’s fundamentals deteriorate its marginal cost of supplying will increase and it will supply less (analogous to liquidity hoarding). Second, if bank j’s fundamentals deteriorate very significantly then it may not be supplied anything (market lockout), because contagion means that supplying bank j is very costly. Third, if common fundamentals deteriorate very significantly then no banks may be supplied anything (market shutdown). Fourth, although risk-sharing is not explicitly in the model (other than in a reduced form sense via $\Gamma_{ij}$), the outcome of our model is similar: in equilibrium links form between banks whose fundamentals are negatively correlated.

Finally, we note that the model is flexible enough to answer the economic questions we’re seeking to answer, in that the magnitude of each of the sources of inefficiency discussed above depends on specific parameters. Market power is determined by $\theta_l$, the efficiency of decentralised cost allocations by $\Gamma_{ij}$ and $X_{it}$, and the extent of negative externalities in network formation depends on $\phi_1$ and $\Gamma_{ij}$. Similarly, the role of regulation depends on the specific parameters: the impact of increasing capital requirements on interbank surplus depends on $\tau$ and equilibrium responses to caps depend, on the demand side at least, on $\theta_l$. In other words, the extent of inefficiency and the effect of regulation are, given our model, an empirical question. We discuss how we answer this empirical question in the next section.

23These drive marginal cost differences across banks, which are the source of this inefficiency.
24If $\tau$ is large then the negative effect of higher $\lambda$ (that is, higher marginal cost) on interbank surplus is dominated by the fact that higher $\lambda$ reduces contagion intensity.
25Product differentiation determines the extent to which demand would be re-allocated to other banks in the event of a cap, which determines the effectiveness of the cap in reducing overall exposures and decreasing risk.
4 Identification

We describe our approach to identifying the network formation and default risk models in detail below, as well as setting out extensions in which we attempt to control for banks that are not within our sample. First, though, we describe in high-level terms the key variation that we use to identify the important parameters in our model.

4.1 Key variation

We need to disentangle the causal effect of network links $C_{ijt}$ on default risk $p_{it}$ from the effect of $p_{it}$ on $C_{ijt}$. We do this by looking for exogenous variation in the fundamentals of a bank, for which we use a weighted-average of regional equity indices where the weights are a bank’s revenues in each region. We use this variable, $Z_{P_{it}}$, as an instrument for $p_{it}$ in the network formation model. We define this more formally below, as well as arguing that it is plausibly exogenous of unobservables in the interbank market when we include the rich set of fixed effects available to us.

To identify the default risk process, we need exogenous variation in $C_{ijt}$ at the pair-level, not just at the bank-level. We do this by using our network formation model, which shows that equilibrium $C_{ijt}$ is a non-linear function of $Z_{P_{it}}$, $Z_{P_{jt}}$, $Z_{P_{k\neq j,i},t}$ and $\lambda_{it}$. We use such non-linear variation, which we denote $Z_{C_{ijt}}$, as an instrumental variable for $C_{ijt}$. One element of this, for example, is $Z_{P_{it}} \times Z_{P_{kt}}$, which reflects the fact that our model shows that $C_{ijt}$ is likely to be high when the fundamentals of bank i are good and those of bank k (which is competing with i to supply bank j) are bad.

Having discussed the basic sources of exogenous variation in our empirical approach, we now describe the three key patterns in the data that identify the important parameters in our model.

1. Comovement between network concentration and capital requirements: One
of the effects of increased capital requirements $\lambda_{it}$ in our model is that they increase the marginal cost of supplying on the interbank market. If bank $i$ responds to this by increasing $C_{ijt}$ relative to $C_{ikt}$ (holding bank fundamentals constant), then this is evidence that the increase in marginal cost was smaller for supplying to bank $j$ than to bank $k$. Within our model, this means that $\Gamma_{ij} < \Gamma_{ik}$. This comovement between network concentration and capital requirements variation helps identify the contagion intensity parameters $\Gamma$.

2. Comovement between network links and default risk: We also obtain identifying information about contagion intensity from the default risk process. If $p_{it}$ is less responsive to $C_{ijt}$ (or rather, to the instrumental variable we have for $C_{ijt}$) than $C_{ikt}$, holding bank fundamentals constant, then this is also evidence that $\Gamma_{ij} < \Gamma_{ik}$. Inter-temporal variation in this comovement identifies $\tau_t$, common time variation in contagion intensity.

3. Comovement between network links supplied to the same bank: Consider bank $i$’s choice of $C_{ijt}$. The fundamentals of bank $k \neq i, j$ do not directly enter bank $i$’s optimality condition for $C_{ijt}$. The fundamentals of bank $k$ do, however, affect $C_{kjt}$, which affect bank $i$’s optimality condition via the interest rate it earns. If $C_{ijt}$ is sensitive to the fundamentals of bank $k$, then this suggests that the products supplied by $i$ and $k$ are close substitutes. This comovement is, therefore, the key source of variation that identifies product differentiation $\theta$. Conditioning this empirical comovement on observable product characteristics then provides information about the role of these characteristics in determining differentiation.

These three key sources of variation identify the most important parameters in our model. It is these parameters that determine (i) the extent of inefficiency in the decentralised equilibrium and (ii) the outcome of our counterfactual analyses.
4.2 Network formation

The EQC allow us to solve for equilibrium $C$ as a function of $p$, $\lambda$ and $X$. $p$ are endogenous, in that they are functions of the unobserved structural error terms. We treat $\lambda$, regulatory capital requirements, as exogenous, in keeping with the literature (Robles-Garcia (2018), Benetton (2018)).

To obtain bank-specific exogenous variation, we take the relevant equity index to be a bank-specific weighted average of global equity indices from S&P, where the weightings are the proportion of the bank’s revenues that come from that geography (data provided annually by Bloomberg, based on corporate accounts). For example, suppose that at time $t$ bank $k$ obtained 70% of its revenues from the US and the remaining 30% from Japan. In this case, $Z_{kt}^p = 0.7 \times S&P500_t + 0.3 \times S&P Japan_t$. Although this is clearly an imperfect measure of the bank’s fundamentals, we argue it has informative value: this bank $k$ would plausibly be more affected by a slowdown in Japan than some other bank with no Japanese revenues.

We therefore use $Z_{kt}^p$ as an instrument for $p_{it}$. Treating this as exogenous assumes that a bank’s revenue distribution and the equity indices themselves are independent of (or at least predetermined relative to) structural errors in the interbank network. We justify the former on the basis that a bank cannot easily shift its revenue base within the required timeframe (HSBC, for example, has deep roots in Asia that cannot easily be moved), nor would it seek to do so in response to the interbank market. We justify the latter on the basis that these banks are small relative to the total values of these equity indices, and that whilst interbank crises can have wider economic impacts (as arguably happened during the crisis), they do not do so contemporaneously within our 3 to 6 month observation frequency. Both of these assumptions are dependent on including $it$ and $jt$ fixed effects: in their absence, endogenous demand and supply outside options would be correlated with such variation.

$B$ is not separately identifiable from the other parameters. We normalise $B = 1$ on the
basis that in models of quantity competition what matters for market power is $\theta/B$, not the absolute value of $B$.

### 4.2.1 Outside options

Some parameters in terms that are not pair-specific are swept up by these fixed effects, including $A_1$ and $A_2$ in demand. If we denote the $jt$ dummies by $\delta_{jt}$:

$$\delta_{jt} = A_1 + A_2 X_{jt} - \theta_0 C_{0jt} + e_{jt}$$

where $e_{jt}$ is a composite error term. Substituting in equilibrium $C_{0jt}$ and $r_{0jt}$:

$$\delta_{jt} = A_1 + [A_2 (1 - \theta_0) + \theta_0 \phi_0 D] X_{jt} - \theta_0^2 \sum_{k \neq 0} C_{kJt} + e_{jt}$$

$C_{kJt}$ are endogenous with respect to $e_{jt}$ but we can use the same instruments as above to recover an estimate of $\theta_0$. That is, although we do not observe the demand outside option, we can recover an estimate of its substitutability with interbank exposures by considering how the $jt$ dummies vary across $j$ and $t$ with respect to observable exposures.

We cannot separately recover estimates of $A_2$ (how a bank’s demand varies with its fundamentals) and $\phi_0$ (how the price of the bank’s outside option varies with its fundamentals), but the distinction between these two parameters is not important for total surplus, so we simply normalise $\phi_0^D = 1$.

### 4.2.2 Incomplete data

As described in Section 2, for non-British banks we only observe local-unit-to-group exposures, under-estimating their total exposure.

We denote local-unit-to-group exposures by $\tilde{C}_{ijt}$ and group-to-group (that is, total) exposures by $C_{ijt}$. We assume that $C_{ijt} = a_i \tilde{C}_{ij}$, where $a_i$ are bank-specific parameters that we
estimate. These parameters \(a\) are identified given that (i) some variables, such as \(X_{jt}\) and \(p_{it}\), enter the EQC with non-bank-specific coefficients and (ii) for the British banks we know \(a = 1\). In principle, \(a_{ijt}\) is identifiable in this way, but we restrict variation to \(a\) to preserve degrees of freedom.

### 4.3 Default risk process

The advantage of explicitly considering network formation is that we can account for the endogeneity of the network in our spatial DRP model.

We obtain pair-specific instruments for \(C_{ijt}\) by noting that equilibrium exposure \(C_{ijt}\) (obtained by solving EQC) is a non-linear function of \(Z^P_{it}, Z^P_{jt}, Z^P_{mt}\) and \(\lambda\). This non-linearity comes from the iterative effects of the competitive process: suppose, for example, that \(Z^P_{it}\) is high (indicating that bank i is exogenously risky) and \(Z^P_{mt}\) is low. There is a first order effect on EQC via \(p_{it}\) and \(p_{mt}\), which is in equilibrium extended through the competitive process: because bank m (a supply competitor to bank i) supplies more to bank j because it is relatively low risk and so has lower cost of equity, bank i supplies less. That is, whilst \(p\) is a linear function of \(X\), the network formation model is such that exposures are a non-linear function of \(X\). In Appendix A we show, for a simplified version of our model for which an analytical solution exists, that equilibrium \(C\) are non-linear function of \(X\).

Non-linear, exogenous variation in bank fundamentals therefore helps us separately identify the effect of \(C\) and the effect of \(p\). We include \(\lambda\) in order to benefit from its exogenous intertemporal variation.

Table 1 also shows that specific relationships are sticky, in that \(ij\) fixed effects explain a large amount of variation in \(C_{ijt}\). We therefore include \(ij\) dummies as instruments, and leverage intertemporal variation in \(\lambda\) by interacting \(\lambda\) with “gravity” style variables designed to capture whatever drives the underlying stickiness in bilateral relationships, including dummy variables indicating (i) \(i\) and \(j\) have their headquarters on the same continent and
(ii) $i$ and $j$ are both legacy investment banks.

We define $\tilde{Z}^P_{ijt} = \frac{1}{N-2} \sum_{k \neq i,j} Z^P_{kt}$ (that is, average fundamentals of other banks) and $g_{ij}$ as these gravity style variables. As instruments for $C_{ijt}$ we use:

$$Z^C_{ijt} = [Z^P_{it}Z^P_{jt}, \tilde{Z}^P_{ijt}Z^P_{jt}, \tilde{Z}^P_{ijt}\tilde{Z}^P_{ijt}, \lambda_{it}Z^P_{it}Z^P_{jt}, \lambda_{it}\tilde{Z}^P_{ijt}Z^P_{jt}, \lambda_{it}\lambda_{it}\tilde{Z}^P_{ijt}, g_{ij}\lambda_{it}]$$

Estimating network effects in a spatial model such as this one further requires that the errors be uncorrelated across banks. That is, positive correlation in unobserved fundamentals would be incorrectly identified as network contagion. We justify this assumption by including $X^2_t$, a set of macroeconomic and commodity variables that we allow to enter with bank-specific coefficients, in a bid to properly account for comovements in fundamentals. That is, the full set of bank fundamentals $X$ in the default risk process consists of $Z^P$ and $X^2$.

### 4.4 Extensions

The network data we rely on is infrequent and does not include every major bank, but the default risk and fundamentals data is available at much higher frequency and for any publicly traded bank. In this sub-section, we describe how we use this to extend our baseline empirical approach in two ways: first, to account for the impact of unobserved banks; and second, to make use of greater inter-temporal variation in the DRP.

#### 4.4.1 Accounting for unobserved nodes

Our baseline empirical approach assumes that the endogenous effect of unobserved links with banks outside of our sample is captured by the relevant fixed effects in the network formation and default risk specifications. We test this by collecting $p$, $X$ and $Z$ for 4 additional, publicly traded banks that are not included in the Bank of England dataset on which we rely. We then do the following, where an “o” (“no”) superscript indicates it relates to one of the
observed (not observed) banks:

• First, we run first-stage regressions as above, by regressing observed exposures on observed instruments.

\[ C_{ijt}^o = Z_{ijt}^o \beta_1^1 + \nu_{ijt}^1 \]

• Second, we use the observed instruments of the unobserved banks and the estimated coefficients from this first stage to calculate exogenous proxies for these unobserved links.

\[ \hat{C}_{ijt}^{no} = Z_{ijt}^{no} \hat{\beta}_1^1 \]

• Third, we take our exogenous estimate of the network as \( \hat{C}^o \) combined with \( \hat{C}^{no} \), and then estimate the network formation and contagion parts of the model in the second state as in the baseline case.

In step 2, we need to use an adjusted version of \( Z \) that excludes bank-specific or pair-specific fixed effects. We show in Table X, and discuss further in the results section below, how these more limited instruments still perform well in the first stage. In step 3, we do not observe the product characteristics of these unobserved exposures; instead we assume constant differentiation for these banks outside of our sample.\(^{26}\)

This extension gives no additional identifying information in estimation of the network formation game, but it accounts for the endogenous effect of these unobserved links. In the case of DRP, it also accounts for endogenous unobserved links, but also gives additional identifying information in estimation, as the default risk of these unobserved banks is observed. It does expand the number of parameters to estimate, as the size of \( \Gamma \) is \( N(N-1) \).

Consider a stylised example that demonstrates the contribution of this approach. Take Bank Z, for which we do not observe any network exposures \( C_{izt} \) or \( C_{zit} \). If these unobserved links are correlated with the fundamentals of other banks in equilibrium and not sufficiently

\(^{26}\)That is, we set \( \theta_{ijt} = \theta_{ijt}^{no} \) if either \( i \) or \( j \) is in the \( no \) set.
homogenous to be captured by the fixed effects, then the instruments described above would likely be invalid. However, we do observe default risk $p_{zt}$ and fundamentals $X_{zt}$. This data, combined with our model, tells us some limited, but useful, information about its likely exposures: for example, we know that $C_{zkt}$ is likely to be higher when $X_{zt}$ is low (indicating its fundamentals are good), $X_{kt}$ is high and $X_{mt}$ is high (where $m \neq z, k$). We can then use the cross-sectional and inter-temporal variation in these proxies to at least partially control for unobserved variation in exposures involving Bank Z in the network formation game and in DRP, preserving the validity of our instruments.

4.4.2 Additional inter-temporal variation

Our network data is collected relatively infrequently (every 3 or 6 months), whereas default risk and fundamentals are traded and so available every trading day. A very similar process to that described above can therefore be used to exploit this additional data: in essence, we regress observed, infrequent exposures on observed, infrequent instruments. We then use the estimated parameters and observed frequent instruments to generate proxies for exposures at the higher frequency. This allows us to test whether our model fits the data at a higher frequency, with significantly more degrees of freedom.

5 Estimation

The parameters we seek to estimate are $\Theta = (\Gamma, \beta, A_1, A_2, \zeta, \tilde{\theta}, \phi_1)$; respectively, contagion intensities, fundamentals, demand intercept and time variation, pairwise technology importance, characteristic-based product differentiation and cost multiplier. We apply GMM jointly on two sets of moment conditions:

- network formation: $\mathbb{E}[Z'(\hat{C}(\Theta) - C)] = 0$; and
- contagion: $\mathbb{E}[Z'(\hat{p}(\Theta) - p)] = 0$. 

31
By way of preliminary estimation, we currently weight the two moment conditions evenly. Possible refinements include: choosing the weighting matrix efficiently and using shrinkage techniques (particularly regarding $\Gamma_{ij}$) to preserve degrees of freedom (following de Paula et al (2018)).

5.1 Results

We set out our results in Tables 5 and 7.

Most importantly, our results indicate that the model fits the data very well. The $R^2$ for network formation is 0.71, indicating that the model has explanatory power beyond the extensive fixed effects used: a regression of $C_{ijt}$ on the relevant fixed effects only has an $R^2$ of 0.52. We set out various F-tests in Table 6, and find that the parameters in our model are collectively significantly different from zero.

We draw the following substantive implications from our results. First, interbank contagion is a significant part of the default risk process: mean $\Gamma_{ij}$ is such that the mean partial equilibrium$^{27}$ elasticity of $p_{it}$ with respect to $p_{jt}$ is around 0.018: that is, for an average exposure $C_{ijt}$, an increase by 1% in $p_{jt}$ causes $p_{it}$ to increase by 0.018%. When aggregated across multiple counterparties and taking into account full equilibrium responses, this effect is material: on average around one quarter of $\bar{p}$ is due to interbank contagion, with the remaining three quarters due to fundamentals.

Second, there is substantial variation in $\Gamma_{ij}$, contagion intensity: certain links are significantly more costly, in terms of the effect on default risk. We plot the estimated distribution of $\Gamma_{ij}$ in Figure 8.

Third, there is evidence that contagion intensity has decreased across our sample, in line with increasing capital requirements. Estimated $\tau$ implies that mean contagion intensity decreased by around 20% between 2011 and 2016. Fourth, there is very limited product

---

$^{27}$That is, without taking into account the fact that an increase in $p_{jt}$ and $p_{it}$, causes other bank’s default risks to increase, which then has a feedback effect on $p_{jt}$ and $p_{it}$. 

substitutability\textsuperscript{28} between the products supplied by banks.

Finally, we find that the decentralised interbank market is not on the efficient frontier: a social planner would be able to increase interbank surplus by 28.8\% without increasing mean bank default risk or decrease mean bank default risk by 12.5\% without decreasing interbank surplus, as set out in Figure 2. This result comes from (i) the fact that we find that supplying banks have market power and (ii) the fact that contagion (and thus network externalities) is significant.

\textbf{Figure 2: Decentralised inefficiency}

\textsuperscript{28}At this stage we only allow for homogenous product differentiation between products, without parameterising with respect to product characteristics. That is, we estimate $\theta$ but restrict $\hat{\theta}_l = 0$.  

33
6 Counterfactual Analysis

Two changes to regulation are currently being considered, as we describe in Section 2: a lower cap on maximum individual exposures and increases to capital ratios. We evaluate the effect of these changes on market outcomes by simulating the new equilibrium using our estimated parameters and assuming that fundamentals are unchanged.

We find that a moderately tighter cap has no impact on $\bar{p}$, because the cap creates excess supply and unmet demand that causes other exposures to increase. That is, the network topology changes endogenously. Given our estimated results indicate there is limited demand substitution, this effect is driven by excess supply. A significantly tighter cap does reduce default risk (as a trivial example the cap could be set at 0, such that there is no contagion), but at the cost of very large decreases in interbank surplus.

A general increase in $\lambda$ across all banks has two effects in our model: first, it increases the marginal cost of supplying financial products on the interbank market; and second, it reduces contagion intensity via parameter $\tau_t$. Our simulations indicate that the former dominates the latter, such that $\bar{p}$ decreases but so does $TS_t$.

We also consider two alternative forms of regulation that, to our knowledge, are not being publicly considered. The first is a cap on total exposures held by bank $i$, rather than on individual exposures. That is, a cap on $\sum_j C_{ijt}$ rather than on $C_{ijt}$. A cap on individual exposures simply induces bank $i$ to supply other banks; a cap on total exposures held by bank $i$ prevents this. Our counterfactual simulations indicate that such a cap does in fact reduce $\bar{p}$, in contrast to a cap on individual exposures, as we show in Figure 3.

The second alternative approach we propose is to take a more targeted approach to $\lambda$, taking into account network externalities. The key parameter in our model is $\Gamma_{ij}$, contagion intensity: links where this is high are particularly costly, in terms of their effect on default risk. We propose a mean-preserving spread of $\lambda$ on a pairwise basis related to contagion intensity. More formally, we divide pairwise links into those where $\Gamma_{ij} < median(\Gamma)$ ("low
Figure 3: Counterfactual analysis of caps
risk links”) and those where $\Gamma_{ij} \geq median(\Gamma)$ (“high risk links”). We then simply increase $\lambda$ by $b$ for high risk links and decrease $\lambda$ by $b$ for low risk links, where $b$ is a parameter governing the size of the spread. Our simulations indicate that this is a Pareto-improvement, in the sense defined above, over current and proposed regulation, as we show in Figure 4.

Figure 4: Counterfactual analysis of capital requirements
7 Conclusion

In this paper we set out a structural model of network formation and contagion. Preliminary empirical results indicate that this model fits the data well and has material policy implications: it enables us to quantify the extent of inefficiency in the interbank market and simulate optimal regulation.

As an extension, we are also considering applications of this model to stress-testing: that is, simulating shocks to bank fundamentals and examining the role of the endogenous interbank network in risk propagation.

References


University Press Cambridge.


Kashyap, A. K., Stein, J. C., and Hanson, S. (2010). An analysis of the impact of substantially heightened capital requirements on large financial institutions. *Booth School of Business, University of Chicago, mimeo, 2*.


## Tables

### Table 1: Variation and persistence in network

<table>
<thead>
<tr>
<th></th>
<th>FE = jt</th>
<th>it</th>
<th>i,j,t</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_{ijt}</td>
<td>R^2 = 0.14</td>
<td>0.18</td>
<td>0.52</td>
</tr>
<tr>
<td>d_{propderiv,t}</td>
<td>0.26</td>
<td>0.24</td>
<td>0.55</td>
</tr>
<tr>
<td>d_{propfi,t}</td>
<td>0.21</td>
<td>0.11</td>
<td>0.41</td>
</tr>
<tr>
<td>d_{propsft,t}</td>
<td>0.25</td>
<td>0.24</td>
<td>0.56</td>
</tr>
<tr>
<td>No. obs</td>
<td>4,590</td>
<td>4,590</td>
<td>4,590</td>
</tr>
</tbody>
</table>

Note: This table shows the R^2 obtained from regressing observed network links or characteristics on dummy variables. jt, for example, indicates that the regressors are dummy variables for each combination of j and t.

### Table 2: Variation and persistence in default risk

<table>
<thead>
<tr>
<th></th>
<th>FE = i</th>
<th>t</th>
<th>i,t</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_{it}</td>
<td>R^2 = 0.17</td>
<td>0.58</td>
<td>0.75</td>
</tr>
<tr>
<td>No. obs</td>
<td>270</td>
<td>270</td>
<td>270</td>
</tr>
</tbody>
</table>

Note: This table shows the R^2 obtained from regressing observed network links or characteristics on dummy variables.
Table 3: First-stage: Network formation results

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
</tr>
<tr>
<td>(a_i)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
</tr>
<tr>
<td>FE</td>
<td>ij, t</td>
<td>ij, jt</td>
<td>ij, it</td>
<td>ij, it, jt</td>
</tr>
<tr>
<td>R(^2)</td>
<td>0.71</td>
<td>0.73</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>No. obs</td>
<td>4,590</td>
<td>4,590</td>
<td>4,590</td>
<td>4,590</td>
</tr>
</tbody>
</table>

Note: We only report coefficient estimates for a subset of the instruments.

Table 4: First-stage: Default risk process

<table>
<thead>
<tr>
<th></th>
<th>(p_{it})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_{it})</td>
<td>0.755*** (r(0.000))</td>
</tr>
<tr>
<td>FE</td>
<td>i,t</td>
</tr>
<tr>
<td>R(^2)</td>
<td>0.85</td>
</tr>
<tr>
<td>No. obs</td>
<td>270</td>
</tr>
</tbody>
</table>

Note: Figure in parentheses is p-value.
Table 5: Network formation results

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
<td>(C_{ijt})</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>0.078</td>
<td>0.077</td>
<td>0.079</td>
<td>0.079</td>
</tr>
<tr>
<td>(\theta)</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
<tr>
<td>(\Gamma_{ij})</td>
<td>0.0088 - 0.038</td>
<td>0.0088 - 0.037</td>
<td>0.0088 - 0.039</td>
<td>0.0088 - 0.038</td>
</tr>
<tr>
<td>FE</td>
<td>(i,j, t)</td>
<td>(i,j, jt)</td>
<td>(i,j, it)</td>
<td>(i,j, it, jt)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.717</td>
<td>0.733</td>
<td>0.736</td>
<td>0.752</td>
</tr>
<tr>
<td>No. obs</td>
<td>4,590</td>
<td>4,590</td>
<td>4,590</td>
<td>4,590</td>
</tr>
</tbody>
</table>

Note: For \(\Gamma_{ij}\) I report the interquartile range. The full distribution is shown in Figure 8.

Table 6: F tests

<table>
<thead>
<tr>
<th>Restricted model</th>
<th>Unrestricted model</th>
<th>F statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects only</td>
<td>Model [1]</td>
<td>16.5</td>
<td>0.000</td>
</tr>
<tr>
<td>Model [1]</td>
<td>Model [2]</td>
<td>0.90</td>
<td>0.859</td>
</tr>
<tr>
<td>Model [1]</td>
<td>Model [4]</td>
<td>0.96</td>
<td>0.726</td>
</tr>
</tbody>
</table>

Note: Fixed effects only includes \(i,j\) and \(t\) fixed effects, to match those included in Model 1.

Table 7: Default risk process results

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(p_{it})</td>
<td>(p_{it})</td>
<td>(p_{it})</td>
<td>(p_{it})</td>
</tr>
<tr>
<td>(\Gamma_{ij})</td>
<td>0.0088 - 0.038</td>
<td>0.0088 - 0.037</td>
<td>0.0088 - 0.039</td>
<td>0.0088 - 0.038</td>
</tr>
<tr>
<td>FE</td>
<td>(i,t)</td>
<td>(i,t)</td>
<td>(i,t)</td>
<td>(i,t)</td>
</tr>
<tr>
<td>(\beta_i)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.910</td>
<td>0.894</td>
<td>0.900</td>
<td>0.883</td>
</tr>
<tr>
<td>No. obs</td>
<td>270</td>
<td>270</td>
<td>270</td>
<td>270</td>
</tr>
</tbody>
</table>

Note: Default risk estimated jointly with network formation. Variations [1] to [4] refer to the controls included in the network formation regression, as set out in Table 5 above.
Figures

Figure 5: Contraction in the interbank market
Figure 6: The interbank market and risk
Figure 7: Consolidation in the interbank market
Figure 8: Histogram of estimated contagion intensity
A Mathematical appendix

A.1 EQC

In this appendix, we derive the equilibrium quantity condition, EQC. The first order supply condition is:

\[ r_{ijt} = -\frac{\partial r_{ijt}}{\partial C_{ijt}} C_{ijt} + m c_{ijt} + \frac{\partial p_{it}}{\partial C_{ijt}} \sum_k \frac{\partial m e_{ikt}}{\partial p_{it}} C_{ikt} - \frac{\partial \Pi_D}{\partial p_{it}} \frac{\partial p_{it}}{\partial C_{ijt}} + r_{i0t} + e_{ijt} \]

It follows immediately from DRP that \( \frac{\partial p_{it}}{\partial C_{ijt}} = \tau_{ij} \Gamma_{ij} \) from our assumed cost function that \( \frac{\partial m e_{kt}}{\partial p_{it}} = \phi_1 \lambda_{ikt} \) and from our demand model that \( \frac{\partial r_{ijt}}{\partial C_{ijt}} = -B \) and \( \frac{\partial \Pi_D}{\partial p_{it}} = -\sum_k \frac{\partial r_{ikt}}{\partial p_{it}} C_{ikt} \).

\[ r_{ijt} = BC_{ijt} + \phi_1 \lambda_{ijt} p_{it} + \phi_1 \tau_{ij} \Gamma_{ij} p_{jt} \sum_m \lambda_{imt} C_{imt} + \frac{\partial p_{it}}{\partial C_{ijt}} \sum_m \frac{\partial r_{mit}}{\partial p_{it}} C_{mit} + r_{i0t} + e_{ijt} \]

For ease of exposition we then repeat the same equation for supply from bank k to bank i:

\[ r_{kit} = BC_{kit} + \phi_1 \lambda_{kit} p_{kt} + \phi_1 \tau_{ki} \Gamma_{ki} p_{jt} \sum_m \lambda_{kmt} C_{kmt} + \frac{\partial p_{kt}}{\partial C_{kit}} \sum_m \frac{\partial r_{mkt}}{\partial p_{kt}} C_{mkt} + r_{k0t} + e_{kit} \]

When bank i considers how much to supply to bank j, it takes into account the impact of the resulting increase in \( p_{it} \) on its profits from being supplied exposures: this is the penultimate term in the equations above. That is, it takes into account the effect of its supply on \( r_{kit} \). We assume that bank i takes the interest rates of transactions involving other parties as given, such that:

\[ \frac{\partial r_{kit}}{\partial p_{it}} = \phi_1 \tau_{ki} \Gamma_{ki} \sum_m \lambda_{kmt} C_{kmt} \]

Substitute this and demand equation [x] into supply, and we obtain the EQC:

\[ 0 = A_1 + A_2 X_{jt} + \zeta_{ij} + e_{ijt}^D - 2BC_{ijt} - \sum_{k \neq i}^N \theta_{ik} C_{kjt} + e_{ijt}^S + \lambda_{ijt} \phi_1 p_{it} - \phi_1 \tau_{ij} \Gamma_{ij} p_{jt} \sum_{k \neq i}^N C_{ikt} \lambda_{ikt} - r_{i0t} - \phi_1 \tau_{ij}^2 \Gamma_{ij} p_{jt} \sum_k C_{kit} \Gamma_{ki} \sum_m C_{kmt} \lambda_{kmt} \]

48
A.2 Equilibrium links are non-linear in fundamentals

Consider a simplified version of the model in which banks do not consider the impact of their supply decisions on \( \Pi_D \); that is, they consider the impact on their funding costs when supplying on the interbank market, but not on their funding costs when demanding from the interbank market. This means that the EQC is linear in \( C \). Furthermore, for simplicity of exposition (and without loss of generality regarding the form of equilibrium \( C \)) suppose \( \zeta = e^D = e^S = r_0 = 0 \), \( 2B = \phi_1 = \lambda = 1 \), \( \theta_{ij} = \theta \), \( \Gamma_{ij} = \Gamma \) for all banks and parameters are such that all equilibrium exposures are strictly positive. The EQC is then as follows:

\[
0 = A_1 + A_2 X_{jt} - C_{ijt} - \theta \sum_{k \neq i}^N C_{kjt} - p_{it} - \Gamma p_{jt} \sum_{k \neq i}^N C_{ikt}
\]

In this case an analytical expression for equilibrium exposures exists, where \( C \) is a \( N(N-1) \times 1 \) vector of endogenous exposures, \( p \) is a \( N \times 1 \) vector of default probabilities, \( X \) is a \( N \times 1 \) vector of fundamentals, \( M_i, M_j, M_{\Sigma i}, \) and \( M_{\Sigma j} \) are matrices that select and sum the appropriate elements in \( C \) and \( p \), and \( \circ \) signify matrix multiplication and the Hadamard product, respectively:

\[
C = \left[ I + \theta M_{\Sigma j} + (M_j.p) \circ M_{\Sigma i} \right]^{-1} \left[ A_1 + A_2 M_j.X - M_i.p \right]
\]

Given that \( p \) is a linear function of \( X \), as set out in the DRP, it follows that equilibrium \( C \) is a non-linear function of \( X \).

B Model simulations

TBC.