Interwoven Lending, Uncertainty, and Liquidity Hoarding*

Adam Zawadowski

Central European University

zawa@ceu.edu

December 13, 2017

Abstract

This paper shows how uncertainties about funding in an interwoven system of intermediation can lead to excessive liquidity hoarding. In the model, funds are channeled through several financial intermediaries (banks) until they are finally invested in real assets. In case of a funding shock, banks that are uncertain about their own loans being rolled over, fear bankruptcy and cannot commit to rolling over loans they made to others either. This fear can lead local funding uncertainties to prompt banks to liquidate inefficiently large amounts of real assets without any defaults in equilibrium. The results hold even though the only source of risk aversion in the model is due to bankruptcy cost, banks are otherwise risk-neutral. The model suggests a novel mechanism for credit crunches observed in crisis episodes.

*I thank Markus Brunnermeier, Alp Simsek, Jeremy Stein, Wei Xiong and participants at the Princeton finance workshop for comments on this and earlier versions of the paper. All remaining errors are mine.

1 Introduction

The system of financial intermediation is interwoven with many layers of intermediaries between the initial investor and the real investment.[1] Also, many of the loans between intermediaries (banks from hereon) are short-term or even overnight and thus can be recalled without any advance notice. I capture this structure by assuming that funds are channeled through a network of intermediaries. In this paper, I show that contagion and liquidity hoarding can be due to funding uncertainty without any defaults in equilibrium.

The key ingredients are the uncertainty due to short-term lending and the interwoven network structure of channeling funds towards real investments. In case of a funding shock, uncertainty is spread and magnified through the network: banks that are uncertain about their own funding cannot themselves commit to rolling over loans they have made, spreading the uncertainty further. The reason that banks hoard cash is that in case their loan is not rolled over, they need to pay back right away: similar to a cash in advance constraint. In equilibrium, banks liquidate inefficient amounts of real assets. Since each bank finances several other banks, uncertainty about funding in one bank spreads to more and more banks in the consecutive layers of intermediation. The inefficiency increases exponentially in the parameters measuring how interwoven the network structure is. This result highlights how the interwoven structure of intermediation magnifies local uncertainties.

Figure 1 shows the way I model the interwoven structure, layering and branching, in the system of financial intermediation. Each bank has a few relationships through which it can channel loans and also has multiple sources of loans to finance its own lending. Also, there are several layers of interwoven intermediation between the initial investors and the real asset. If the layering of interwoven intermediation can have such devastating effects in crisis, why do initial investors not directly lend to firms that do the real investments? There are two potential driving forces (not modeled in the paper). First, lending can only be sustained among institutions that trust or can monitor each other (Karlan et al. 2009). Second, layering and branching helps achieve diversification. While data is not available for the whole financial system, Bech and Atalay (2008) analyzes the federal funds market and finds that funds lent on average go through another 3.5 banks before reaching their final destination. This

[1] See Bech and Atalay (2008) for evidence of such structure in case of interbank loans.

yields indirect evidence for the existence of several layers of intermediation between initial investors and banks who use the funds.

The results also help understand disruption in interbank lending during the Financial Crisis that started in August 2007. Models concentrating on asset side contagion cannot give a full explanation for the difficulties in obtaining funds and liquidity hoarding unless one is willing to assume some kind of extreme, e.g. Knightian, risk aversion. The reason is that, at least during the beginning of the crisis, ex post default rates could not explain the high interest rates on 3 month interbank loans as measured by LIBOR (Armantier et al. 2008). Thus a model in which there is liquidity hoarding without defaults, such as this one, helps us understand the disruptions in interbank markets in the absence of cascading defaults in the financial system. Also, in this paper, banks have simple linear utility with bankruptcy cost: no Knightian risk aversion is needed to generate the results.

The model can also be used to understand the importance of a public liquidity backstop in case of funding uncertainty, such as the Term Auction Facility implemented early in the Financial Crisis. In the model, such a liquidity backstop can stop uncertainty from spreading through the system and prevent excessive liquidation. I also show why such a liquidity backstop cannot be provided by the banks themselves. Given the network externality through the funding links, banks do not find it privately optimal to establish such a facility: much of the benefits accrue to banks who do not contribute to it.

The paper is structured as follows. Section 2 relates the findings to previous literature. Section 3 gives the simple baseline model and illustrates the main mechanism. Section 4 discusses how public provision of liquidity can overcome inefficiency and why private arrangements to provide liquidity are not feasible. Finally, Section 5 concludes and discusses implications for future research.

2 Related literature

In traditional models of financial networks, contagion spreads from borrower to lender through a domino of defaults, such as Kiyotaki and Moore (1997) and Allen and Gale (2000). This paper proposes a fundamentally different explanation for contagion: I show that banks might hoard liquidity because they are uncertain about their own funding, i.e. crisis spreads through risks on the liability side of the balance sheet. Zawadowski (2010) also models contagion through the liability side: a systemic run of creditors in a network of OTC hedging contracts that can lead to the collapse of the whole system. However, there is a major difference: in this model no banks fail in equilibrium. This is important since Furfine (2003) and Upper and Worm (2004) show empirically that cascading defaults through the asset side of the balance sheet are unlikely to cause large contagion and a systemic crisis. Thus the mechanism proposed in this paper gives an alternative explanation for excessive liquidity hoarding that does not rely on cascading defaults or Knightian uncertainty.

There are several other papers of networks in economics that are related to my results. The forward looking nature of recalling investments is similar to that in Lagunoff and Schreft (2001), who develop a highly stylized model of links in the financial system. In a dynamic model they also find that banks preemptively cut links after losses to reduce their risk exposure, however, their results are hard to interpret in the context of credit. Bak et al. (1992) have a simple mechanical model of inventories and production in an economy. The structure of production is reminiscent of the interwoven intermediation network studied in this paper: they get multiplication because every firm's output is the input of two others. However, they get amplification in mechanical way, not through strategic decisions prompted by uncertainty like in this paper. Caballero and Simsek (2009) extend the model of Allen and Gale (2000) to show that the threat of asset side contagion leads to liquidity hoarding if there is uncertainty about the network structure. On the other hand, their results rely on a cascade of defaults and Knightian risk aversion.

The paper also relates to dynamic inefficiency in time. He and Xiong (2009) shows how dynamic refinancing of short-term loans can lead to an inefficient outcome since lenders to not internalize their decision's effect on lenders whose debt matures later. Acharya et al. (2009) shows that refinancing of short-term debt can be inefficient depending on the sequential realization of information. In this paper, the inefficiency is not due to inefficiency over multiple periods in time as in the previous papers but due to inefficiency over multiple layers of intermediation.

The model of the paper can most readily be applied to interbank markets. Armantier et al. (2008) discusses the disruptions in term lending between banks during the Financial Crisis of 2007-2008. They also argue that lending activity dropped in the term funding market both because of worries about creditworthiness of counterparties and lenders worries about their own ability to raise funds in the future. This latter channel is the one modeled in this paper. The model of this paper also helps understand why the Federal Reserve's Term Auction Facility might have helped ease the stress of term lending in the interbank market.

3 Baseline model

3.1 Model setup

The model consists of three periods: $t = 0, 1, 2, 3$. In the financial system, there are $L \geq 2$ layers of financial intermediaries (banks from hereon). Figure 1 presents an example for a part of the system. I leave the linkages in the whole system unmodeled and choose to concentrate on small part for expositional simplicity: in reality there might be loops in the systems, etc. The layers are referred to as A,B,C,... where the top layer is A, the second from top is B, etc. The top layer of the banks are the initial sources of funds, each of them endowed with K (integer) units of cash at $t = 0$. Each bank in layer A has links enabling it to transfer one unit of cash each to K banks in the next layer B. Each bank in layer B receives cash from K different banks in layer A. The number K is the branching number of the network. Similarly, banks in the next layer receive cash from the banks in the previous layer. Note that all banks intermediate K loans irrespective of their position in the network. The banks in the last L^{th} layer can invest cash in K real assets. Each link represents a relationship in the network that can support a loan of size unity. I assume for now that no lending can take place by side-stepping these links, the loans are not collateralized, and are short term, i.e. they can be recalled at $t = 1$ without any further notice. This could be an efficient outcome given an underlying moral hazard problem (Calomiris and Kahn 1991, Diamond and Rajan 2001) or an inefficient maturity rat race equilibrium such as modeled in Brunnermeier and Oehmke (2009).

The real assets have a maximum size of one unit but the investment size can be chosen continuously in $[0, 1]$. The investment in the real asset has to be made at $t = 0$. If held to maturity at $t = 3$, each real asset yields $1 + R$ per unit of investment. A continuous fraction of the real asset can be liquidated at $t = 1$, yielding one unit of cash per unit of initial investment liquidated. Cash can be stored costlessly.

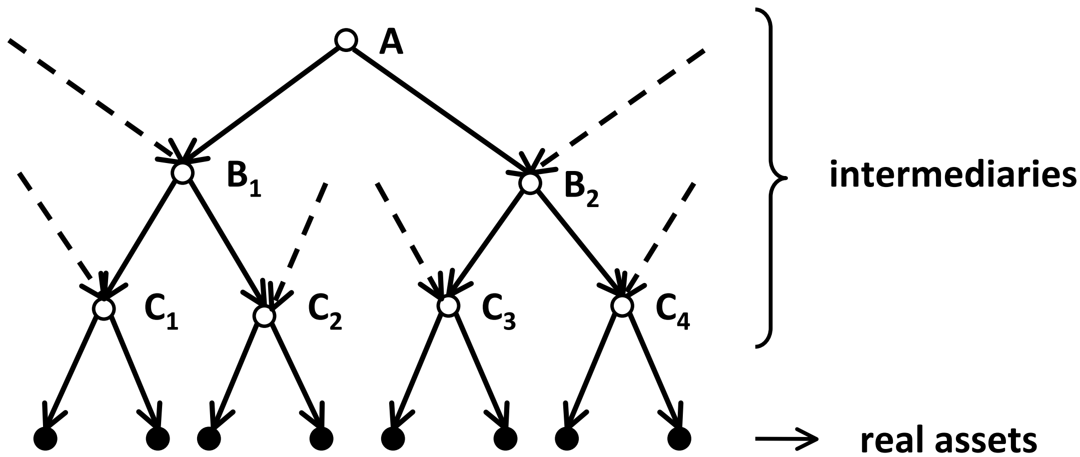

Figure 1: Interwoven lending

Network structure with $L = 3$ layers and branching number of $K = 2$. The hollow circles are intermediaries, the solid ones real assets. The arrows show the direction of loans, the dashed arrows are the loans that come from outside the analyzed subsection of the network.

Each link, if utilized to give a loan, requires a fixed monitoring cost of $m = \frac{M}{L}$ irrespective of the size of the loan; the same holds for investing in the real asset. I scale the costs with L in order to make a conservative assumption: overall monitoring cost per dollar invested does not change with the number of layers L. The fixed cost does not have to be paid if the loan is completely recalled at $t = 1$. Each bank receives $r = \frac{R}{L}$ commission, or interest rate differential, per unit of loan it extends. I assume $r > m$, thus it is worth it for banks to extend loans. A bank goes bankrupt if at $t = 1$ its loan is recalled and it does not have the cash to repay it: thus there is a cash in advance constraint. For simplicity, a bankrupt institution is completely liquidated (i.e. none of its loans get rolled over) and due to bankruptcy costs all of its lenders receive a recovery value of only $1 - b$ per unit of loans, where $b > 0$. If a real project is liquidated at $t = 1$, there is a possibility to reinvest ex post, once the crisis has ebbed, at $t = 2$. However, the real return on such a project is lower, $R' < R$, and each bank along the intermediation chain receives a commission of $r' = \frac{R'}{L}$. I assume that reinvestment is still worthwhile, i.e. $r' > m$.

Banks receive payoffs due to intermediation and realize utility at $t = 3$. They are risk-neutral, but since payoff is conditional on survival, this makes them behave as if they were risk-averse at $t = 1$.

If banks directly or indirectly lend to both projects with return R and projects with return R', the commission is shared evenly between lenders, i.e. every bank gets a commission based on the average return of the real projects it channels funds to.

At $t = 1$ a specific bank at the level A, to which for simplicity I refer to as bank A, receives an idiosyncratic funding shock: the amount of funds it can lend out decreases by $s \in [0, 1]$, which is a random variable with cumulative distribution function of $F(s)$. Due to the interwoven structure of lending, this single bank A is indirectly financing K^L real assets. The value of s is realized at $t = 1$ before the decision of banks whether to roll over debt. I allow for the distribution $F(s)$ to be partially or completely discrete, e.g. for $s = 0$ to happen with a positive probability.

All banks can decide whether to roll over their loans at $t = 1$, however, they cannot credibly signal the banks they are lending to about which loans they are intending to recall and which they are planning to roll over. All banks decide whether to roll over loans simultaneously and thus cannot condition their decisions on the actions of the other banks. I rule out cooperation between the banks: this seems like a reasonable assumption to model real life crises, especially ones that are fast. All banks know the network structure. For the lending decision I use the equilibrium concept of pairwise stability from Jackson and Wolinsky (1996): loans are established if they are mutually beneficial but can be terminated by either party unilaterally. If a loan is terminated by the borrower at $t = 1$ and the previous lender has available funds at $t = 2$, it randomly chooses which of its non-utilized links to use for new lending at $t = 2$.

3.2 Model solution

3.2.1 The proposed equilibrium

The following three parameter restrictions are imposed to make the problem interesting. Both the monitoring cost and the bankruptcy cost are assumed to be non-trivial compared to loss from liquidating unnecessarily at $t = 1$, i.e. $m > \frac{K-1}{K} \cdot (r - r')$ and $b > \frac{K}{K-1} \cdot (r - r')$. Also, the return net of monitoring costs on reinvesting at $t = 2$ is assumed not to be very low compared to investing at $t = 0$: $r' - m > \frac{r-m}{K+1}$.

I propose the following subgame perfect equilibrium, given the above parameter restrictions. There exists an s^*, such that if the shortfall is small $s \leq s^*$, bank A simply scales back all its K loans by $\frac{s}{K}$, thus spreading the shortfall evenly. All other banks getting loans from A spread the shortfall evenly too, thus in the end all real assets getting indirect funding from A are scaled back by $\frac{s}{K^L}$. On the other hand, if the shortfall is larger $s > s^*$, then there is inefficient liquidation: all banks getting direct or indirect funding from A deny to roll over one of their loans chosen at random. The banks on the lowest level all liquidate one of their real investments, thus in the end the amount of real assets liquidated is $K^{L-1} \gg 1$ when $s > s^*$, instead of just $s < 1$ when $s \leq s^*$. Banks basically liquidate one of their loans each for fear that their lenders, getting loans from A directly or indirectly, also do so. I call this event a crisis and the ensuing behavior of the banks is referred to as a panic. Thus after the potential loan recalls, unused liquidity is stuck in the system (see Figure 2) which then has to be reinvested at $t = 2$, albeit at a lower return. For now let me ignore the possibility that banks hold surplus cash or provide liquidity insurance to each other, I explore this case in Section 4.

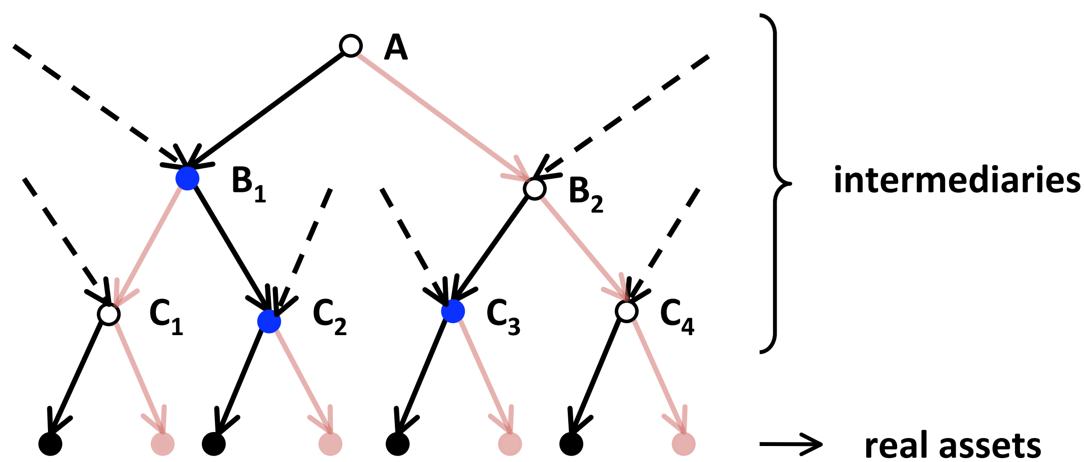

Figure 2: Excessive liquidation

The figure depicts lending ex post after loan recalls but before reinvestment. Liquidity trapped on intermediary balance sheets in the network is depicted by filled circles. Recalled loans and liquidated real projects are lightly shaded, surviving loans and real assets are black. Network structure with $L = 3$ layers and branching number of $K = 2$.

3.2.2 The proof for large shortfall: crisis

The proof is as follows. Let us first analyze the case $s > s^*$ and I show that the proposed equilibrium indeed exists (see Figure 2 for a realization of loan callbacks). When bank A makes a decision about recalling a full loan, it rationally anticipates that the inefficient liquidation that follows will decrease its expected payoff. At $t = 1$ bank A recalls one of its K loans and does not take into account payoffs along this branch any more (the right branch in Figure 2). Bank A rolls over $K - 1$ of its loans (the left branch in Figure 2), however, given the crisis and excessive liquidation, in equilibrium $\frac{1}{K}$ portion of real projects using financing from these loans are liquidated but then reinvested in $t = 2$ for return R'. This means that the average payoff on these loans decreases from R to $\frac{K-1}{K} \cdot R + \frac{1}{K} \cdot R'$. Bank A recalls one loan completely if and only if the average payoff with the recall is higher than that when it does not completely recall a loan but just reduces loans by a total of s. It also takes into account that by recalling a full loan, instead of simply scaling back loans, it saves m on monitoring costs since it has one loan less to monitor. Recalling a full loan is privately optimal for bank A if and only if:

$$(K-1) \cdot \left[\frac{K-1}{K} \cdot r + \frac{1}{K} \cdot r'\right] - (K-1) \cdot m > (K-s) \cdot r - K \cdot m \tag{1}$$

thus a full loan is recalled if $s > s^*$, where:

$$s^* = 1 - \frac{m}{r} + \frac{K-1}{K} \cdot \frac{r-r'}{r} \tag{2}$$

where the second term reflects that the higher the monitoring cost, the more likely a full loan recall is. The third term is an adjustment for the effect that bank A knows that recalling a loan will lead to a crisis and make the average return on the real assets lower. This leads bank A to be more conservative in order to avoid a crisis and excessive liquidation and thus it recalls a full loan only when s is high. To make sure that $s^* < 1$, i.e. crisis is possible for some $s \in [s^*, 1]$, we assumed that the monitoring cost is not very small: $m > \frac{K-1}{K} \cdot (r - r')$.

Banks along the chain of intermediation B, C, etc. know that if they recall a loan they can reestablish it at $t = 2$. However, they (except for the last bank financing the real project directly) also rationally anticipate that calling back a loan leads to a crisis and thus the average return on real assets decreases. They also know that if a bank lending to them recalls one of its loans, with probability $\frac{1}{K}$ this loan is the one extended to them. In case the bank lending to them does not roll over the loan extended to them and they do not have the cash on hand to pay it back, they go bankrupt and get payoff 0. Note that since they already face a financing shortfall of one full unit, not only $s < 1$, these banks always choose to liquidate a full loan if bank A chose to. Thus they recall one of their loans if and only if:

$$\frac{K-1}{K} \cdot K \cdot \left[\frac{K-1}{K} \cdot (r-m) + \frac{1}{K} \cdot (r'-m)\right] + \frac{1}{K} \cdot (K-1) \cdot \left[\frac{K-1}{K} \cdot (r-m) + \frac{1}{K} \cdot (r'-m)\right]$$

$$> \frac{K-1}{K} \cdot K \cdot (r-m) + \frac{1}{K} \cdot 0 \tag{3}$$

The first part of the left hand side is the expected payoff in case the bank has liquidated one of its loans but in the end the loan it received did not get recalled (like ending up in the situation of bank B_1 in Figure 2) which occurs with probability $\frac{K-1}{K}$. In this case at $t = 2$ all of the bank's K loans get rolled over after temporary recall of a single loan at $t = 1$. The second term is the expected payoff of the bank if it recalled one of its loans and also one of its own loans was recalled (like ending up in the situation of bank B_2 in Figure 2). This event occurs with probability $\frac{1}{K}$ and leaves the bank with $K - 1$ loans. It follows that the condition for recall is:

$$r' - m > \frac{r-m}{K+1} \tag{4}$$

which holds unless r' is very low, i.e. unless the profit per unit of intermediation on a post-crisis investment is less than $\frac{1}{K+1}$ of the profit on a unit of pre-crisis investment. This assumption could be relaxed by adding some additional continuation value conditional on survival at $t = 1$. The bank that finances the real projects directly faces a slightly different decision, since it does not have to face the problem that others that it extends financing to will recall loans. This means that on the left hand side of Equation 3 the expected payoff from loans when recalling a full loan is r. Since $r > \frac{K-1}{K} \cdot (r-m) + \frac{1}{K} \cdot (r'-m)$ it follows that whenever the intermediate banks recall a full loan, so will the last bank that finances the real projects directly. Thus we have shown that if bank A recalls one loan, all other banks that directly or indirectly get funding from bank A also recall one loan each, chosen at random.

There are two other potential deviations from this equilibrium strategy to analyze before concluding that the proposed equilibrium exists. First, banks may choose to hold cash ex ante to avoid having to liquidate a project: I explore this possibility on Section 4.1. Second, banks may terminate one of their borrowing relationships if they know they will recall one of their loans. If they do not return funds they have borrowed, i.e. stick to the loan, there is no additional cost. On the other hand, in case a bank chooses to give up one unit of funding at $t = 1$, it might happen that another bank gets to have the loan at $t = 2$ if it turns out that funding is available after all. This is due to the assumption that when reestablishing loans at $t = 2$, the lending bank randomly chooses which of its non-utilized links it uses for new lending.

3.2.3 The proof for small shortfall: no crisis

Now let us analyze the case $s < s^*$, that is when the funding shortfall is less than the critical amount. In this case it is an equilibrium that bank A splits the reduction in project size between its K projects, each project is downsized by $\frac{s}{K}$. Also, it is an equilibrium that all other intermediaries split the downsizing equally between their projects. In the end this leads to all real projects, receiving direct or indirect funding from bank A, being reduced by $\frac{s}{K^L}$. I choose to focus on this equilibrium since it is symmetric and socially efficient, even though it is not unique. There is a continuum of equilibria: banks do not necessarily have to choose to scale back all loans by the same amount. However, all other strategies lead to some inefficient liquidation since the banks that A is lending to do not know which loan will be scaled back by the most and this uncertainty would lead to liquidity hoarding as in the case of $s > s^*$.

There is a potential deviation from the strategy of scaling back all loans by the same amount: withdraw a full loan even if $s < s^*$, when banks that bank A has lent to only expect a scaling back of loans. In this case the bank, the loan of which is not rolled over, goes bankrupt since it did not itself call back loans and bank A only receives the face value of the loan net of bankruptcy costs: $1 - b$.

This deviation is not chosen if scaling back loans is more profitable:

$$(K-1)\cdot r-(K-1)\cdot m-b<(K-s)\cdot r-K\cdot m \quad (5)$$

the left hand side of the equation is the payoff from the deviation, taking into account that other banks do not anticipate it and thus do not liquidate real projects and that the bank can save on monitoring costs. This also means that bank A loses b due to bankruptcy if it still insists on not rolling over the loan. The right hand side is the payoff from simply scaling back loans without fully recalling any of them. Substituting $s<s^*$ (from Equation 2) we arrive at a lower bound for b to ensure the deviation is not profitable:

$$b>\frac{K-1}{K}\cdot(r-r') \quad (6)$$

and this is what we have assumed about the borrowing cost. The reason that the borrowing cost has to satisfy the same lower bound as the monitoring cost is that of bankruptcy was not costly enough then bank A would want to call on the loans of banks at level B by surprise.

3.3 Comparison to benchmark without interwoven lending

Let us compare the results to a benchmark case without interwoven intermediation, which corresponds to the previous model with $L=1$. In this case all banks at level A have direct access to K final projects at an intermediation fee of $M=L\cdot m$. Bank A collects all of the return $R=L\cdot r$ on the real projects. Using this scaling ensures that the total economic cost of monitoring is the same in both structures, with and without the interwoven structure. In this case with direct financing, bank A chooses to fully liquidate a real project instead of downsizing all loans if and only if:

$$(K-1)\cdot R-(K-1)\cdot M>(K-s)\cdot R-K\cdot M \quad (7)$$

the left hand side of the equation now does not take into account that panic reduces the expected payoff of real projects, only that recalling a full loan economizes on monitoring costs. This leads to a cutoff value s^{**}: if at $t=1$ bank A is hit by a funding shock of $s<s^{**}$, it scales back all projects by $\frac{s}{K}$, a total of s. There are also other equilibria which do not entail symmetric downsizing but

these are not discussed since they have exactly the same real effect. However, if $s > s^{**}$, bank A fully terminates a real project. The cutoff value is:

$$s^{**} = 1 - \frac{m}{r} \tag{8}$$

Note that this cutoff is lower than that for the interwoven system $s^{**} < s^*$. Figure 3 compares the amount of liquidation with and without the interwoven structure of lending as a function of the amount of initial shortfall s.

There is a welfare loss due to the interwoven structure of intermediation and the potential crisis, see Figure 4. The amount liquidated without an interwoven structure is the socially optimal amount since there is no uncertainty due to which loans will not be rolled over; the figure shows the welfare loss in the interwoven system compared to this benchmark. The amount of liquidation K^{L-1} in the interwoven system is inefficient if $s > s^*$ since the efficient amount of liquidation would be only one unit. Given that $K^{L-1} - 1$ projects are liquidated unnecessarily but the reinvested at $t = 2$, the overall amount of welfare loss is: $(K^{L-1} - 1) \cdot (R - R')$. Inefficiency results from the amplification due to uncertainty and crisis in the interwoven intermediation structure. Furthermore, the amount of overall liquidation and thus the inefficiency increases exponentially in the branching number and the number of layers. Thus a slightly more complicated system of intermediation can result in a substantially larger amount of liquidity hoarding in crisis.

There is also welfare loss if $s \in [s^{**}, s^*]$ since bank A is afraid of calling back a full loan even though it would be socially optimal to do so. The reason is that the prospect of such a recall sets off a crisis and lowers expected payoff for bank A even on loans that are not recalled. In this interval loans are only scaled back, instead of one being recalled completely. The social payoff from scaling back a loan is $(1 - s) \cdot R - M$, while that from recalling a full loan would be 0. Thus the welfare loss is $M - (1 - s) \cdot R$, which takes value of 0 at s^{**} and $\frac{K-1}{K \cdot (R-R')}$ at s^*. If $s < s^{**}$ there is no welfare loss, since both with and without intermediation bank A only chooses to scale back projects.

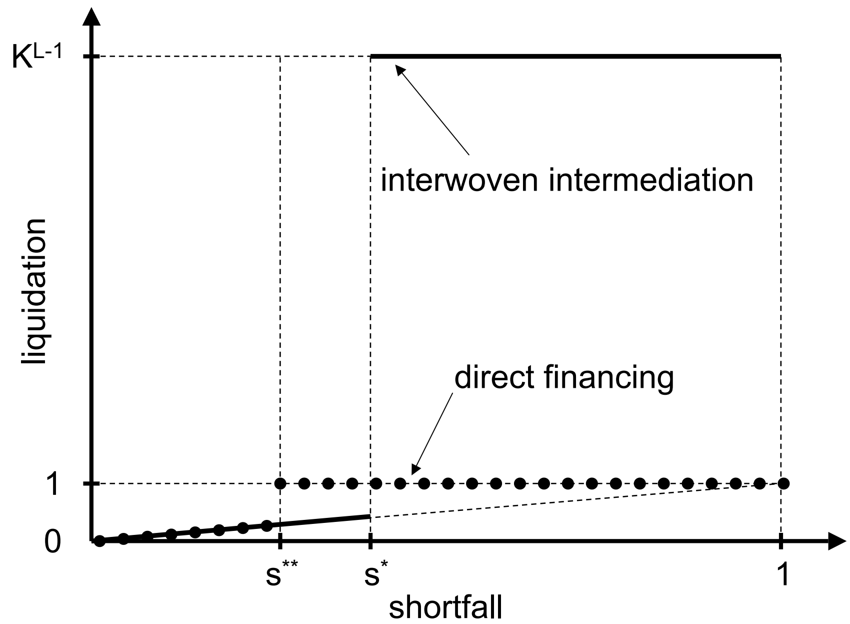

Figure 3: Liquidation as a function of funding shortfall

The solid line is the amount of real liquidation at $t = 1$ with interwoven intermediation, while the line with the circles is the amount of liquidation in the benchmark with direct lending.

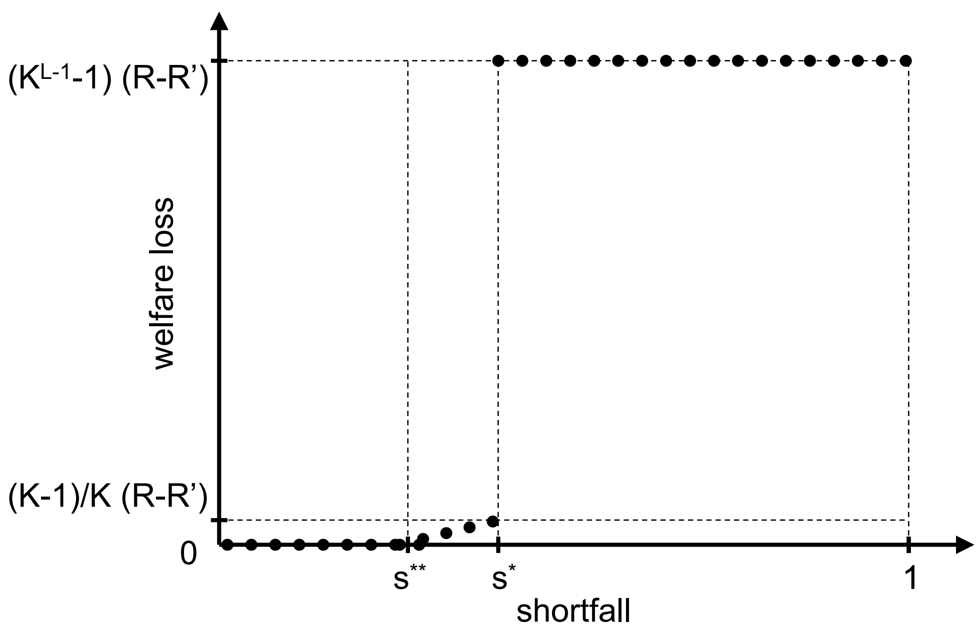

Figure 4: Welfare loss as a function of funding shortfall

The line with the circles is the difference between the welfare in the model without interwoven intermediation and that with the interwoven structure.

3.4 Discussion of model assumptions

One of the assumptions driving the results of the paper is that banks randomly choose which loan not to roll over. This assumption is realistic if banks have several loans of similar quality and e.g. wait until the last minute for some kind of news realization that helps them pick which of the loans to terminate. Alternatively they might have several loans, the quality of which they cannot tell apart, and indeed choose at random. One might wonder why banks do not have clear pecking orders about which of their loans they are going to terminate first. What would be needed to overcome the uncertainty driving the crisis in this paper would be a list of the pecking order of loans, publicly available and enforceable: i.e. a bank would not be allowed to recall loans in any other order than that specified in this list. This seems unrealistic, since it would entail making lots of private information public about beliefs in loan quality and would also curb the ability of the bank to use the threat of termination inherent in short-term loans as a monitoring device (Calomiris and Kahn 1991, Diamond and Rajan 2001). Note that having different priority groups of loans does not overturn the results. Even if there is a group of loans that the bank commits to not withdrawing, e.g. by making them longer term, the uncertainty about which loan is going to be recalled is still present within the group of loans with lower priority.

Another crucial assumption is the timing of loan callbacks. The results depend on the assumption that once the loan extended to a bank is not rolled over it does not have the possibility to recall another loan it has extended. Basically there is a cash in advance constraint: if a loan is recalled, the bank has to pay it back right away. While this assumption might seem extreme in the long run, it is a reasonable assumption in the short run, in the time span a crisis unfolds. Usually the acute panic phase of a crisis is very short, few days or weeks; this is the time frame in which I believe such an assumption is realistic.

The returns on the real project are assumed to be shared equally along the chain of intermediation. I use this assumption to avoid having to deal with bargaining along the intermediation chain which is not the main focus of this paper.[2] This assumption means that if there is a crisis then the amount of payoff for each intermediary gets reset. Given that financing is assumed to be short-term this seems like a reasonable assumption. If instead we assumed that the intermediation fees are fixed unless a

[2]For a model that takes into account network formation when competing for intermediation rents, see Goyal and Vega-Redondo (2007).

loan is recalled, the main results would still hold. The only qualitative difference would be that the inefficiency region in $[s^{**}, s^*]$ disappears, since bank A does not take into account that by withdrawing a loan, it unleashes a crisis. The reason is that if the intermediation fee that A gets does not depend on the actions of the banks it lends too, it does not take into account whether they liquidate in panic or not.

4 Public and private liquidity provision

In this subsection, I analyze whether banks choose to self-insure against liquidity shocks or set up some kind of private arrangement of credit lines. I show that network externalities might prevent banks from achieving the social optimum. I then discuss what social policies are needed to maximize social welfare. For simplicity let me restrict the distribution of initial funding shortfall $F(s)$ to be bimodal:[3] in the bad state with probability p the funding shortfall is $s = s_b$ while in the good state with probability $1 - p$ there is no shortfall thus $s = 0$. To make the problem interesting and simple, I make two assumptions. First, that the monitoring cost is less than half of the real returns $\frac{m}{r} \leq 0.5$.[4] Second, that having a funding shock is "rare", at least in the weak sense that it is less likely than not having a shock $p < 0.5$. Since the return net of monitoring cost on the real investment is $R - M = L \cdot (r - m)$, I take this as the opportunity cost of holding cash.

In analyzing the provision of insurance, I set aside the issue of how loans received from the liquidity pool are monitored. While this is indeed a strong assumption, it also gives a conservative estimate of the parameter regions in which liquidity insurance arrangements, especially private ones, are sustainable. If one believes that there are serious issues related with monitoring such liquidity provision, then liquidity insurance is even less likely to be sustainable and the inefficiencies due to the interwoven intermediation structure are even stronger.

[3] This assumption is only made to illustrate the point of the paper in a simple way. If we had a more complicated distribution for s one would have to take into account strategies in which cash reserves only partially cover the shortfall.

[4] If one thinks of hedge funds that are perceived as having high costs, they charge 2% of assets and 20% of returns. Given an average annual return of 10-20% before fees, this is still less than half of the total return.

4.1 Holding cash and private liquidity insurance

Holding cash is a way of self-insuring against funding shortfall. In the following, I study when it is optimal for bank A to set aside cash for the case that it experiences a funding shortfall.

There are two cases to consider: if $s_b < s^*$ then the funding shock does not set off a crisis even if bank A does not have any cash reserves, while if $s_b \geq s^*$ it does. In the first case, if the bad state only leads to scaling back of investments, i.e. $s_b < s^*$, then bank A will hold cash reserves of s_b if and only if the benefit is at least as large as the costs. The only benefit in this case is that bank A can hold on to all its investments and collect the return r even on the part s_b that it otherwise would have to scale back:

$$p \cdot s_b \cdot r > (R - M) \cdot s_b \tag{9}$$

where the left hand side is the expected benefit of holding cash of s_b, while the right hand side is the cost of cash. Thus banks A holds cash if and only if the funding shock is frequent, i.e. it is above the cutoff value of $p > \bar{p}^{private}_{s_b<s^*}$, where:

$$\bar{p}^{private}_{s_b<s^*} = \frac{R - M}{r} = \left(1 - \frac{m}{r}\right) \cdot L \tag{10}$$

Given our assumptions that $\frac{m}{r} = \frac{M}{R} \leq 0.5$ and that the system is interwoven, i.e. the number of layers is at least two $L \geq 2$, then the implied cutoff is larger than one. This means that bank A would never set aside cash to overcome a shortfall that simply entails scaling back investments.

If the shortfall is large enough to set off a crisis, i.e. $s_b \geq s^*$, then bank A could still choose to only scale back investments like above. Since bank A finds it privately optimal to withdraw a full loan instead of scaling all loans back, the above calculation gives a lower bound on the probability p when bank A wants to hold cash in order to avoid calling back a full loan when hit by the funding shock in the bad state. As argued above, with realistic parameter values even this necessary condition is unlikely to be met. Alternatively, in the case $s_b \geq s^*$, bank A might choose to use a mixture of the two options: both hold cash and scale back loans somewhat. However, in the case that $p < \bar{p}^{private}_{s_b<s^*}$, this would never be optimal since any incremental cash holding would only result in extra expected payoff of $p \cdot r$ per unit of cash; similarly to Equation 9, the bank would always want to decrease its

cash holdings. Once it reached the point in which it has so little cash that it is indifferent between liquidating a full project or scaling back, the bank simply chooses to liquidate a full loan and decrease the cash holdings to zero. Thus strategies consisting of a mix of both loan scale-backs and cash reserves cannot be optimal, the only strategies are the corner solutions of full cash coverage or no cash, i.e. the ones analyzed above.

Banks in layer B face a similar decision about holding cash and the above derivation applies to them. However, in case the funding shortfall is large, i.e. $s_b > s^*$, there is a choice which is superior to simply holding cash: create a liquidity pool among themselves and then whichever bank's loan is not rolled over can get a loan supplied by the liquidity pool. There are K banks on level B that receive financing from bank A and they have to provide $\frac{1}{K}$ units of cash each to this pool. The benefit is that in the bad state, i.e. with probability p, they do not have to worry about their loan not being rolled over; the cost is the opportunity cost of holding cash. Banks will only join the liquidity pool if and only if:

$$\frac{1}{K} \cdot (R - M) < p \cdot \left[K \cdot (r - m) - \frac{(K+1) \cdot (K-1)^2}{K^2} \cdot (r - m) - \frac{(K+1) \cdot (K-1)}{K^2} \cdot (r' - m) \right] \quad (11)$$

where the right hand side is the difference of $K \cdot (r - m)$, the payoff in crisis if a bank has liquidity insurance, and the left hand side of Equation 3, which is the payoff to a bank in layer B in case there is a crisis. Thus the necessary and sufficient condition for this insurance arrangement is that the probability of the bad state is larger than a cutoff value $p > \bar{p}^{private}_{s_b > s^*}$, where:

$$\bar{p}^{private}_{s_b > s^*} = \frac{K \cdot L}{(K^2 + K - 1) - (K^2 - 1) \cdot \frac{r' - m}{r - m}} = \frac{L}{1 + \frac{K^2 - 1}{K} \cdot \frac{r - r'}{r - m}} > \frac{L}{K + 1} \quad (12)$$

which by dropping the negative terms in the denominator is reduced to the necessary condition on the right hand side. Thus if the shortfall is not very frequent, i.e. $p < \bar{p}^{private}_{s_b > s^*}$, then the liquidity insurance scheme breaks down because no bank on level B is willing to contribute to the liquidity pool.

4.2 Public liquidity insurance: policy implications

In the model, if the shortfall s in bank A is observed by a social planner, such as the central bank, then simply announcing that banks can get a loan from the central bank can stop the panic and the crisis if $s > s^*$. The reason is that if e.g. banks in layer B know that bank A can go to the central bank if it experiences a shortfall, there is no need for them to preemptively withdraw one of their own loans. Thus a credible announcement itself is enough to stop the crisis. In Subsection 4.3, I discuss how this relates to policy interventions during the Financial Crisis of 2007-2008.

In this section, I derive when it is socially optimal for the central bank to create such a liquidity backstop. Clearly it is the most efficient for the central planner to give a liquidity backstop to bank A and thus stop the crisis right at the origin. First I study the case in which $s_b < s^*$ and then if $s_b > s^*$. It is worth for the social planner to set aside cash to avert a scaling back of loans if and only if:

$$(R - M) \cdot s < p \cdot R \cdot s \tag{13}$$

thus the probability of the funding shock has to be large enough $p > \bar{p}^{public}_{s_b<s^*}$, where:

$$\bar{p}^{public}_{s_b<s^*} = 1 - \frac{m}{r} \tag{14}$$

Clearly $\bar{p}^{public}_{s_b<s^*} < \bar{p}^{private}_{s_b<s^*}$ thus there is a region of inefficiency. However, with realistic parameters as assumed, i.e. if $\frac{m}{r} \leq 0.5$, then a funding shortfall has to be more likely than no shock at all in order to be in the inefficiency region. Thus I do not consider this case to be interesting from a policy perspective.

The more interesting case is when the shortfall is large $s_b > s^*$ since bank A's action can then prompt a crisis with panic and massive liquidation. Given the simple assumptions that $\frac{m}{r} \leq 0.5$ and $p < 0.5$ by which it is never even socially optimal to avert the scaling back of projects (by Equation 14), the social planner just has to make sure that bank A does not find it optimal to liquidate a full loan. The least costly way of doing this is by giving bank A a free loan of $s_b - s^*$, just enough to make it indifferent between scaling back and full liquidation. Clearly the most important part of the policy is for it to be public and credible, s.t. banks at level B can be sure about their loans being rolled over, even if scaled back, and do not panic and spread the funding uncertainty. The social cost again is the cost of the government holding cash of $s_b - s^*$ but the benefit is foregoing unnecessary liquidation of $K^{L-1} - s^*$ real projects. Note that there is still orderly liquidation of real projects up to a total amount of s^*. Thus public provision of liquidity to bank A is socially optimal if and only if:

$$(R - M) \cdot (s_b - s^*) < p \cdot \left[(K^{L-1} - s^*) \cdot (R - R') + (R - M)\right] \tag{15}$$

since K^{L-1} is a large number in general, it implies that already for very low probability events the central planner would want to provide a liquidity backstop to bank A, more specifically, if and only if $p > \bar{p}^{public}_{s_b > s^*}$, where:

$$\bar{p}^{public}_{s_b > s^*} = \frac{s_b - s^*}{K^{L-1} - s^*} \cdot \frac{r - r'}{r - m} \tag{16}$$

where s^* is defined by Equation 2.

One can easily show that it always holds that $\bar{p}^{public}_{s_b > s^*} < \bar{p}^{private}_{s_b > s^*}$, thus there is a range of probabilities $p \in [\bar{p}^{public}_{s_b > s^*} < \bar{p}^{private}_{s_b > s^*}]$, s.t. the private optimum is not socially optimal.[5] Furthermore, as the system becomes more and more interwoven, i.e. more layers and branching, the term K^{L-1} grows exponentially and $\bar{p}^{public}_{s_b > s^*} \rightarrow 0$. That is even for very unlikely crises the central bank would find it socially optimal to provide the above liquidity backstop. Thus the results of the baseline model hold up even if liquidity insurance is possible. The reason is that the externalities still prevent the banks from setting up socially optimal liquidity insurance facilities ex ante.

4.3 The mechanism and the Financial Crisis of 2007-2008

The insights of the model in this paper help give a rationale for interventions such as the Term Auction Facility (TAF from hereon). The Federal Reserve implemented the TAF in December 2007 due to disruptions in interbank lending. The intervention gave banks access to loans against a wide range of collateral. The intervention was coordinated with several other central banks, including the European Central Bank and the Bank of England. Armantier et al. (2008) show that TAF auctions were well and widely utilized by banks. McAndrews et al. (2008) shows that the announcement of TAF did indeed

[5] $\bar{p}^{public}_{s_b > s^*} < \bar{p}^{private}_{s_b > s^*}$ follows from the following observations: $\bar{p}^{private}_{s_b > s^*} > \frac{L}{K+1}$; $L \geq 2$; $\frac{r-r'}{r-m} < 1$; $s_b < 1$.

decrease the LIBOR-OIS spread, which is a measure of how hard it is for banks to get term funding: 1 and 3 month unsecured loans. Their study also shows evidence that it is the announcements and not the auctions that ease lending pressure. This finding is in line with the implications of the model, since if banks in layer B know that they will have access to funding from the Federal Reserve in case bank A does not roll over their loan, they will be willing to roll over the funding they provide for banks in layer C.

While not directly modeled, this paper gives insights regarding the question why banks resorted to overnight lending instead of term loans during the crisis. If there were enough funds in the short run, i.e. for a few days but uncertainty about the funds available in the longer run, the collapse of lending in the model can be applied to term loans. Banks are much less willing to give out term loans if they are uncertain about their own term loans being rolled over, they will resort to giving out overnight loans, which they can do given that the funding uncertainty is not imminent.

5 Conclusions

This paper shows how channeling short-term funds through an interwoven system of intermediation might cause local funding shocks to result in massive liquidation of real investments. These results do not rely on the extreme assumption of Knightian uncertainty: bank have linear utility with bankruptcy cost. This privately optimal liquidity hoarding is socially suboptimal. Banks choose not to hold a cash buffer; thereby uncertainty about their own funding creates further funding uncertainty for the banks they lend to: this is the network externality that leads to an inefficient outcome. The results thus show how complicated interwoven financial intermediation can be destabilizing and how public liquidity backstops can prevent panic and unnecessary liquidation. This paper also shows that banks and financial intermediaries cannot set up such backstops privately, unless they are compelled to.

Given the results, future theoretical research should focus on better understanding the effects of financial networks. This paper shows that the behavior of network-based intermediation can be very different from the purely market-based system. Empirical research is also needed to map out the network structure and the types of links within financial system in order to better understand which network effects are likely to be of practical importance. Understanding these effects is not only theoretically interesting but is important from a policy perspective, in order to devise better financial regulation.

References

Acharya, Viral, Douglas Gale, and Tanju Yorulmazer, "Rollover Risk and Market Freezes," *New York University, Working Paper*, 2009.

Allen, Franklin and Douglas Gale, "Financial Contagion," *Journal of Political Economy*, 2000, *108* (1), 1–33.

Armantier, Olivier, Sandy Krieger, and James McAndrews, "The Federal Reserve's Term Auction Facility," *Current Issues in Economics and Finance*, 2008, *14* (5), 1–10.

Bak, Peter, Kan Chen, José A. Scheinkman, and Michael D. Woodford, "Aggregate Fluctuations from Independent Sectoral Shocks: Self-Organized Criticality in a Model of Production and Inventory Dynamics," *NBER Working Paper, No. 4241*, December 1992.

Bech, Morten L. and Enghin Atalay, "The Topology of the Federal Funds Market," *Federal Reserve Bank of New York, Staff Report No. 354*, 2008.

Brunnermeier, Markus and Martin Oehmke, "The Maturity Rat Race," *Princeton University, Working Paper*, 2009.

Caballero, Ricardo J. and Alp Simsek, "Complexity and Financial Panics," *NBER Working Paper, No. 14997*, 2009.

Calomiris, Charles W. and Charles M. Kahn, "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review*, 1991, *81* (3), 497–513.

Diamond, Douglas W. and Raghuram G. Rajan, "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, 2001, *109* (2), 287–327.

Furfine, Craig, "Interbank Exposures: Quantifying the Risk of Contagion," *American Economic Review Papers and Proceedings*, 2003, *35* (1), 111–128.

Goyal, Sanjeev and Fernando Vega-Redondo, "Structural holes in social networks," *Journal of Economic Theory*, 2007, *137* (1), 460 – 492.

He, Zhiguo and Wei Xiong, "Dynamic Debt Runs," *Princeton University, Working Paper*, 2009.

Jackson, Matthew O. and Asher Wolinsky, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 1996, *71*, 44–74.

Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl, "Trust and Social Collateral," *Quarterly Journal of Economics*, 2009, *124* (3), 1307–1361.

Kiyotaki, Nobuhiro and John Moore, "Credit Chains," *Princeton University, Working Paper*, 1997.

Lagunoff, Roger and Stacey L. Schreft, "A Model of Financial Fragility," *Journal of Economic Theory*, 2001, *99*, 220–264.

McAndrews, James, Asani Sarkar, and Zhenyu Wang, "The Effect of the Term Auction Facility on the London Inter-Bank Offered Rate," *Federal Reserve Bank of New York, Staff Report No. 335*, 2008.

Upper, Christian and Andreas Worm, "Estimating bilateral exposures in the German interbank market: Is there a danger of contagion?," *European Economic Review*, 2004, *48*, 827–849.

Zawadowski, Adam, "Entangled Financial Systems," *Princeton University, Working Paper*, 2010.