

Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics^{*†}

Amil Dasgupta
London School of Economics

Abstract

Financial contagion is modeled as an equilibrium phenomenon in a dynamic setting with incomplete information and multiple banks. The equilibrium probability of bank failure is uniquely determined. We explore how the cross holding of deposits motivated by imperfectly correlated regional liquidity shocks can lead to contagious effects conditional on the failure of a financial institution. We show that contagious bank failure occurs with positive probability in the unique equilibrium of the economy and demonstrate that the presence of such contagion risk can prevent banks from perfectly insuring each other against liquidity shocks via the cross-holding of deposits. (JEL: G2, C7)

^{*}Acknowledgements: I am grateful to the editors, Franklin Allen and Patrick Bolton, and to two anonymous referees for detailed and helpful comments. This paper is a revised version of a chapter of my Ph.D. dissertation at Yale University. I would like to thank my advisor, Stephen Morris, and committee members, Ben Polak and Dirk Bergemann for their guidance. This paper has benefited from discussions with V. V. Chari, Itay Goldstein, Timothy Guinnane, Patrick Kehoe, Jonathan Levin, John Moore, Ady Pauzner, Debraj Ray, Andreas Roider, and Hyun Shin. I thank Elu von Thadden for his insightful discussion of this paper at the CFS Conference on Liquidity Concepts and Financial Instabilities, 2003, and seminar participants at LBS, LSE, the Bank of England, and Yale, for comments. Financial support from the Yale's Cowles Foundation and Northwestern's CMS-EMS is gratefully acknowledged. All remaining errors are my own.

[†]Email address: a.dasgupta@lse.ac.uk

1 Introduction

A commonly held view of financial crises is that they begin locally, in some region, country, or institution, and subsequently “spread” elsewhere. This process of spread is often referred to as *contagion*. What might justify contagion in a rational economy? There are (at least) two broad classes of explanations.

The first class of explanations posits that the adverse information that precipitates a crisis in one institution also implies adverse information about the other. This view emphasizes correlations in underlying value across institutions and Bayes learning by rational agents.¹ For example, a currency crisis in Thailand may be driven by adverse information about underlying asset values in South East Asia, which can then apply to other countries in the region.

A second type of explanation begins with the observation that financial institutions are often linked to each other through direct portfolio or balance sheet connections. For example, entrepreneurs are linked to capitalists through credit relationships; banks are known to hold interbank deposits. While such balance sheet connections may seem to be desirable *ex ante*, during a crisis the failure of one institution can have direct negative payoff effects upon stakeholders of institutions with which it is linked.²

In this paper, we present a model of financial contagion which formalizes this latter view. We focus on a particular (but particularly important) type of financial institution: commercial banks. Throughout history, banks have cross-held deposits (for clearance, regulatory and insurance reasons), and thus the failure of some banks had direct consequences on others through capital linkages. Contagious bank failure is particularly complex because it involves an underlying coordination problem amongst depositors of each bank. Even weak banks may not fail if very few depositors withdraw their money early, while strong banks may fail if many depositors withdraw early. The existence of multiple equilibria (Diamond and Dybvig 1983) makes it difficult to examine even individual bank failures, which then compounds the difficulty of isolating contagious effects in many bank settings. Using and extending some recent developments in the theory of equilibrium selection in coordination games (Morris and Shin 2003), we present a model of an economy with multiple banks where the probability of failure of individual banks, and of systemic crises, is uniquely determined. This then permits us to identify contagion precisely and examine its properties.

The model presented here is too stylized to be a realistic depiction of any particular set of financial crises. While it may have some stylized similarities to aspects of episodes both in history (e.g. the pre-World War I panics of the National Banking Era in the United

¹See, for example, Chen (1999) or Acharya and Yorulmazer (2002).

²See, for example, Kiyotaki and Moore (1997) or Allen and Gale (2000).

States) and in current times (e.g. recent episodes of volatility in Latin America and East Asia), our purpose is to provide further theoretical grounding for the existence of financial contagion in equilibrium, and to show that it may occur with positive probability in banking systems. To motivate the model, however, it is useful at the outset to briefly describe the broad stylized features of the former group of financial crises mentioned above.

The defining characteristics of the National Banking System were laid out in the National Banking Act of 1864. This act prohibited interstate branching of banks and established a system of reserve pyramiding, under which country banks could hold reserves in designated reserve city banks, which in turn could hold reserves in New York. Thus, throughout this period, the reserve cities including New York directly or indirectly held the deposits of many country banks.

There were five major banking panics of in the National Banking Era prior to the Great Depression. They occurred in 1873, 1884, 1890, 1893, and 1907. With the exception of 1893, these panics began in New York and subsequently spread to the interior of the country. The work of Calomiris and Gorton (1991), Wicker (2000) and others indicate that the panics typically began with local shocks to assets in New York. This was typically followed by suspension of payments by New York banks, followed by suspensions in banks at various parts of the country. In 1907, for example, the panic began due to an unsuccessful attempt to corner the Copper market by a group of speculators who were associated with several Trust Companies in New York. When news of this speculative failure became public in October there were runs on Knickerbocker Trust Company. This was followed by runs on the National Bank of North America and on other institutions thought to be directly or indirectly linked to the Copper speculators, and then by a widespread panic. Sprague (1910, p. 259) points out: “Everywhere the banks suddenly found themselves confronted with demands for money by frightened depositors . . . Country banks drew money from city banks and all banks throughout the country demanded the return of funds deposited or on loan in New York.” Finally, the panic that began with a localized asset shock in New York led to suspensions through much of the country.

To summarize, two very broad stylized features of the National Banking System panics were as follows:

- Panics originated due to local asset-side shocks. They were inherently dynamic, starting in New York and spreading to the interior of the country.
- While other factors may also play a role, panics appeared to diffuse nationally through the correspondent network, from debtor New York banks to creditor banks in the interior.

Both of these stylized features emerge as equilibrium outcomes in our model, which we now proceed to describe.

1.1 Summary of Model and Results

We consider an economy with three periods and two regions. Each region has a representative bank. Each bank has access to a (common) liquid riskless asset and a local illiquid risky asset. The risky assets pay a higher expected return than the riskless asset if held to maturity, but less than par value if liquidated early. The returns on the risky assets are revealed in the last period, and are increasing functions of regional economic fundamentals.

There are two groups of risk-averse depositors, one in each region, each of whom lives three periods. The depositors receive uninsurable private liquidity shocks: they may need to consume in the interim period with positive probability. The total level of liquidity demand in the economy is fixed, but there may be (negatively correlated) regionally aggregate liquidity fluctuations in the interim period. The two banks insure against such regional liquidity shocks by exchanging deposits in the initial period. Consumer and interbank deposits take the form of demand deposit contracts.

In the interim period, regional liquidity shocks are realized first and become publicly known. The bank facing high liquidity demand withdraws its interbank deposit. This creates an interim asymmetry amongst the two banks: one bank is now a net debtor to the other. Later in the same period, the depositors of a randomly chosen bank receive some information about the state of fundamentals in their region. This updates their beliefs about eventual returns on bank deposits and thus leads them to consider whether to leave their deposits in the bank or to withdraw. This may result in runs on that bank. The depositors of the other bank are able to observe the proportion of the customers of the first bank who withdraw their deposits prematurely. They are then able to obtain information about the fundamentals of their own region and choose whether to withdraw their deposits from their own bank.

As we have noted above, the choices made by depositors at each bank involve a coordination problem: they may wish to withdraw their deposits if they believe that enough other depositors will do the same, leading to a run on the bank. This leads to multiple equilibria when payoffs to depositors are common knowledge. In our setting, however, the information available about economic fundamentals in either region is imperfect. Depositors receive private but correlated signals about the potential future returns on their deposits. The level of correlation can be arbitrarily high. However, even small asymmetries in the information received by depositors can lead to substantial *strategic uncertainty*, i.e., uncertainty about the actions of other depositors (who condition their behavior on their own private signals). The

presence of such strategic uncertainty prevents depositors from coordinating their actions with arbitrary precision, and thus greatly reduces the set of potential equilibrium outcomes for a given level of economic fundamentals. Accordingly, in contrast to the usual multiplicity of equilibria that arise in the classic bank runs models in the tradition of Diamond and Dybvig (1983), the game between our depositors is characterized by a unique threshold in regional economic fundamentals below which each bank will fail (Proposition 1) due to a run by depositors.³ Bank failure thus depends upon the release of adverse information about local asset returns. The probability of failure is determined endogenously.

Given the interim debtor-creditor relationship between the regional banks, the emergence of adverse information about asset returns in one region can have broader consequences, causing instability in other regions. In our central result, we show that contagion arises in equilibrium: that is, there are regions of fundamentals in which one bank fails if and only if the other bank fails (Proposition 2). Conditional on the failure of a debtor bank, a creditor bank fails for a wider range of its own fundamentals than if the debtor bank survived. Contagion thus flows from debtors to creditors, and spreads along the channels of interbank deposits. We also present a comparative statics result to demonstrate that the incidence of contagion is increasing in the size of interbank deposit holdings (Proposition 3).

Interbank deposits thus enable banks to hedge regional liquidity shocks, but expose them to the risk of contagion. Should banks hold interbank deposits? Our framework enables us to consider the optimal level of interconnectedness within banking systems. We illustrate the conditions under which banks would want to hold maximal levels of interbank deposits. More importantly, when bank runs are relatively frequent, we show that only *partial* cross-holdings of deposits may be optimal (Remark 1). Thus, the existence of contagion prevents banks from insuring each other perfectly against liquidity risk via the cross-holding of deposits. When the probability of bank failure is high banks may also find it optimal to hold excess reserves as liquidity buffers against depositor runs. Our model suggest that such unstable banking systems may be characterized by lower levels of optimal interconnectedness and higher excess liquidity buffers compared to their stable counterparts. Thus it is precisely in stable banking systems that the rare event of bank failure induces the most significant contagious consequences (Remark 2).

1.2 Related Literature

Our paper is connected with a diverse literature. We apply the equilibrium selection techniques summarized in Morris and Shin (2003). These are discussed further in Section 3.

³The equilibrium selection mechanism and the role of strategic uncertainty is discussed in detail in Section 3.

Goldstein and Pauzner (2000a) were the first to apply these techniques to the analysis of bank runs. They investigate the probability of bank runs in a single-bank setting, while we are interested in the problem of contagion with multiple banks. Rochet and Vives (2000) also analyze bank runs using similar techniques, but do not concern themselves with the problem of contagion. Goldstein and Pauzner (2000b), like us, examine contagion of self-fulfilling crises, but their mechanism for contagion, through common lenders, is different from ours. A related mechanism, based upon a wealth effect, can be found in Kyle and Xiong (2001). Lagunoff and Schreft (2001) study contagion in an economy where the set of available projects, each with a minimum funding requirement, is characterized by overlapping groups of common lenders. When idiosyncratic shocks close down a given project, lenders to that project may optimally reallocate their portfolios. This, in turn, affects lenders who shared other projects with them. Kiyotaki and Moore (1997) explore the method by which contagion flows through credit chains amongst lenders and entrepreneurs. Their model shares with ours the feature that balance-sheet connections are the channels for contagion, but does not concern itself with coordination problems. Rochet and Tirole (1996) examine correlated bank failures via monitoring: the failure of one bank is assumed to mean that other banks have not been monitored, and thus triggers multiple collapses. Freixas, Parigi, and Rochet (1999) consider the role of a lender of last resort in a complete information framework where the presence of systemic risk arises from interbank connections. These connections are motivated by the fact that consumers are uncertain about where they need to consume. Fragility arises from the fact that there may be multiple self-fulfilling equilibrium actions for consumers at each location.

The paper that is closest to ours in theme is by Allen and Gale (2000). Their purpose is to model contagion as an equilibrium phenomenon in a many-bank setting. While our model thus shares features with Allen and Gale's, there are important differences. Contagion is assumed to occur with zero probability in Allen and Gale's model. Thus they do not consider the optimal systemic level of interbank deposit holdings. We are able to determine the probability of contagion in equilibrium and thus can show that full insurance against liquidity shocks via interbank deposits is not always optimal. In addition, our models differ in the source of bank failure. In Allen and Gale's work the source for bank failure is excess liquidity demand. In our model banks fail due to local shocks to bank assets which generate runs by depositors.

The rest of the paper is organized as follows. In the next section we present the model. In section 3 we prove the existence and uniqueness of equilibrium. Section 4 contains our central result and the related comparative static. The optimal level of interbank deposit holdings is illustrated in Section 5. Section 6 concludes.

2 The Model

2.1 Regional Liquidity Shocks

We consider an economy with two non-overlapping “regions,” A and B . There are three time periods $t = 0, 1, 2$. The regions are populated by distinct continuums of weakly risk averse agents with utility functions $u(\cdot)$ [$u'(\cdot) > 0$, $u''(\cdot) \leq 0$] who each live for three periods. Each agent has an endowment of 1 unit. Agents face private (uninsurable) liquidity shocks: they need either to consume in period 1 (impatient) or in period 2 (patient). In the aggregate, there is no uncertainty about liquidity in the economy: there is exactly a proportion $w \in (0, 1)$ of agents who require early liquidity. However, individual regions experience (regionally) aggregate liquidity shocks of size $x > 0$. In particular, there are two states of the world: $\lambda = A$ or $\lambda = B$, corresponding to the cases where region A and region B have high early liquidity demands respectively. Since aggregate liquidity is constant, regional liquidity shocks are negatively correlated. The state λ is realized and publicly known

	A	B
$\lambda = A$	$w + x$	$w - x$
$\lambda = B$	$w - x$	$w + x$

Table 1: Regional Liquidity Shocks

immediately at the beginning of period 1. States A and B occur with equal probability.

2.2 Banks, Demand Deposits, and Interbank Insurance

We consider two representative competitive banks which lie in two regions of the economy. There are two classes of assets: a safe and liquid storage technology with a low (unit) gross rate of return, and a risky, illiquid asset with high expected return but with costs to premature liquidation. The storage technology is common to both banks. One unit stored at time t produces one unit at time $t + 1$. In addition, region i 's bank also has access to risky illiquid technology R_i , with returns given by:

$$R_i(t) = \begin{cases} r \in (\underline{r}, 1) & \text{when } t = 1, \\ R(\theta_i) & \text{when } t = 2, \text{ where } \theta_i \text{ is distributed uniform on } [L, U] \end{cases}$$

where t is the time of liquidation, $R(\cdot)$ is any increasing function, and $\underline{r} \geq \frac{1}{2}$. The parameter θ_i indexes some underlying “fundamentals” related to the bank’s assets, which determine the level of the bank’s asset returns. These fundamentals θ_i are independent and identically

distributed for $i = A, B$. We assume that $E_{\theta_i}[u(R(\theta_i))] \geq u(1)$, i.e., the risky asset pays a higher expected return if held till period 2. Agents cannot directly invest in the risky asset, and begin their lives with their endowments deposited in the bank of their region.⁴

Banks are constrained to offer depositors *demand deposit contracts*. Demand deposit contracts offer conversion of deposits into cash at par on demand in period 1 conditional on sufficient cash being available. If, however, sufficient cash is not available, then the contract specifies that the bank will divide up evenly what cash it can generate by liquidating its portfolio amongst the depositors who demand early withdrawal. At this point of time, the bank goes out of business. For those depositors who choose to remain in the bank till period 2, the bank promises to pay a stochastic amount, which is contingent upon the returns on the bank's assets, the proportion of early withdrawals, and payouts to any senior liabilities.

The two banks face aggregate demand shocks in period 1, in keeping with the regionally aggregate liquidity shocks outlined above. However, since these aggregate regional liquidity shocks are negatively correlated, banks insure against these by holding interbank deposits. In particular, we assume that banks hold cash reserves equal to w , the *average* level of liquidity demand in the economy, and insure against regional liquidity shocks by holding interbank deposits of size $D \in [0, x]$ with the other bank. Given the cash holdings of the banks, and given the timing of the model, it is easy to see that interbank deposits of size larger than x will not be desirable to banks. Thus, in this symmetric scheme, banks exchange deposits of size D , and distribute their net wealth of 1, putting w in cash, and $1 - w$ in long term investment projects. We thus fix banks' portfolios ex ante. However, we relax this restriction below in Section 5.

At the beginning of period 1, the state A or B is realized, and the bank in the high liquidity demand region receives a payment of D from the bank in the other region (before individual depositors can claim money from the bank). In period 2, the debtor bank must pay both its own residual customers and its interbank claim to the creditor bank. We assume that the creditor bank is paid first, and patient depositors share the remainder. The assumed priority order for clearing at $t = 2$ is innocuous: giving interbank payments priority minimizes contagion at the cost of increasing the probability of bank runs in debtor institutions. Since the goal of this exercise is to show that contagion is an essential element of interconnected banking systems, this assumption actually works against us. Changing the relative seniority of interbank debt would lead to qualitatively similar results.

To fix ideas, it is helpful to consider an example. Suppose that only impatient agents withdraw money in period 1 and that $D = x$. In addition, suppose state A is realized, so

⁴We are thus not considering participation. However, in the examples computed below, it is easy to see that participation in the banking system is indeed optimal.

that region A has a higher immediate liquidity shock. Upon the realization of the state, bank A immediately receives from bank B its deposit of x , so that bank A now has $w + x$ in cash, which matches the amount of withdrawals it faces. Similarly, bank B now has $w - x$ in cash, which is precisely the demand it faces in period 1. Bank A now owes bank B the amount $xR(\theta_A)$, and owes its own customers $(1 - w - x)R(\theta_A)$. But it has exactly $(1 - w)$ invested in the illiquid asset $R(\theta_A)$, so its proceeds in period 2 are $(1 - w)R(\theta_A)$, which is exactly the sum of its liabilities. Similarly, promises and earnings clear for bank B .

2.3 Information and Timing

In period 1 nature selects at random (and with equal probability) one of the sets of depositors to receive information about their bank and to act. Information is received in the form of private signals about the underlying fundamentals of their bank. Suppose region i is selected first. Depositor j of region i receives signal $\theta_{j,i} = \theta_i + \epsilon_{j,i}$, where $\epsilon_{j,i}$ are distributed uniformly in the population on $[-\epsilon, \epsilon]$. Shortly thereafter, the depositors of the other bank (in region $-i$) receive information about their own bank, and get to act themselves. The information structure is symmetric. Depositor j of region $-i$ receives signal $\theta_{j,-i} = \theta_{-i} + \epsilon_{j,-i}$, where $\epsilon_{j,-i}$ are distributed uniformly in the population on $[-\epsilon, \epsilon]$. Importantly, before choosing, the depositors who move second learn what happened in the first bank. Thus, the timing of this game can be described below in itemized form:

- Period 0: Interbank deposits are initiated.
- Period 1
 - State A or B is realized.
 - Period 1 interbank claims settle.
 - Depositors in bank i receive information and choose actions.
 - Depositors of bank i who demand early withdrawal are paid.
 - Depositors in bank $-i$ receive information and choose actions.
 - Depositors in bank $-i$ who demand early withdrawal are paid.
- Period 2
 - Period 2 interbank claims settle.
 - Residual depositor claims on the two banks settle.

2.4 Depositor Payoffs and Interbank Payments

We are now ready to write down the payoffs to depositors in this game. In period 1, depositors choose whether to demand conversion of their deposits into cash at par (withdraw) or to retain their deposits with the bank (remain). Impatient depositors can only consume in period 1. They will therefore always withdraw. However, the patient depositors face a non-trivial decision problem. We explicate their payoffs below.

Recall that in period 1 one bank will be a debtor and one bank will be a creditor. Thus, without loss of generality, we can label the payoff matrices for the patient depositors of the two banks as those of the debtor bank and the creditor bank respectively.

Begin by considering the debtor bank, i.e. the bank that experienced a high liquidity shock in period 1. There is a mass $1 - (w + x)$ of patient agents in the debtor region. Let n_d represent the proportion of the patient depositors who choose to withdraw in period 1. If n_d proportion of patient depositors withdraw, then, since impatient agents (of measure $w + x$) always withdraw in period 1, total demand for cash in period 1 is $(w + x) + n_d(1 - (w + x))$. The bank had w in cash and received D in cash from the creditor bank at the beginning of period 1 (and hence became a debtor to the creditor bank). Thus, its total cash holdings are $w + D$. If demand for cash exceeds $w + D$, the bank can obtain more cash by liquidating its long assets. It has $1 - w$ invested in the long asset, from which it can generate $(1 - w)r$ in cash in period 1. Thus, observe that if $[w + x] + (1 - [w + x])n_d \geq [w + D] + (1 - w)r$, i.e., if

$$n_d \geq \frac{(1 - w)r + D - x}{1 - (w + x)} \quad (1)$$

then the debtor bank becomes insolvent and goes out of business in period 1, and in the process divides up the proceeds of its liquidated asset portfolio equally amongst its claimants in period 1. However, if the bank remains solvent in period 1, then it must first settle its debt of $DR(\theta_i)$ to the creditor bank (because interbank deposits have seniority, within each period, to regular demand deposits). In order to pay early demands by patient agents in period 1, the debtor bank had to liquidate $\frac{(1-w-x)n_d+(x-D)}{r}$ of the illiquid asset in period 1. Its original investment in the long asset was $1 - w$. The remaining proceeds are $(1 - w - \frac{(1-(w+x))n_d+(x-D)}{r})R(\theta_i)$. As long as $(1 - w - \frac{(1-(w+x))n_d+(x-D)}{r})R(\theta_i) > DR(\theta_i)$ (i.e., $n_d < \frac{(1-w)r+(D-x)-rD}{1-w-x}$), the debtor bank pays $DR(\theta_i)$ to the creditor bank in period 2, and divides up the remainder equally amongst its residual depositors who chose to remain in the bank. This means that each patient depositor who chooses to remain receives $\frac{1-w-\frac{(1-(w+x))n_d+(x-D)}{r}-D}{(1-w-x)(1-n_d)}R(\theta_i)$. However, if $n_d \geq \frac{(1-w)r+(D-x)-rD}{1-w-x}$, residual depositors receive nothing, and the creditor bank receives (due to seniority) $(1 - w - \frac{(1-(w+x))n_d+(x-D)}{r})R(\theta_i)$. Thus, the period 2 payments on the interbank deposits from the

debtor to the creditor bank can be written as:

$$g(\theta_i, n_d) = \begin{cases} DR(\theta_i) & \text{if } n_d < \frac{(1-w)r+(D-x)-rD}{1-w-x} \\ (1-w - \frac{(1-(w+x))n_d+(x-D)}{r})R(\theta_i) & \text{if } \frac{(1-w)r+(D-x)-rD}{1-w-x} \leq n_d < \frac{(1-w)r+(D-x)}{1-(w+x)} \\ 0 & \text{if } n_d \geq \frac{(1-w)r+(D-x)}{1-(w+x)} \end{cases}$$

Correspondingly, the payoffs to the patient depositors, if they withdraw, are given by:

$$u_W(\theta_i, n_d) = \begin{cases} u[1] & \text{if } n_d < \frac{(1-w)r+(D-x)}{1-(w+x)} \\ u[\frac{(w+D)+(1-w)r}{(w+x)+(1-(w+x))n_d}] & \text{if } n_d \geq \frac{(1-w)r+(D-x)}{1-(w+x)} \end{cases}$$

And if they remain:

$$u_R(\theta_i, n_d) = \begin{cases} u[\frac{1-w - \frac{(1-(w+x))n_d+(x-D)}{r} - D}{(1-w-x)(1-n_d)}R(\theta_i)] & \text{if } n_d < \frac{(1-w)r+(D-x)-rD}{1-w-x} \\ u[0] & \text{if } n_d \geq \frac{(1-w)r+(D-x)-rD}{1-w-x} \end{cases}$$

The specific form of $t = 2$ payments to depositors is a consequence of the assumption of perfectly competitive banks which make zero profits. This is a convenient simplification, but is not necessary for the results below. Given the liquidation discount on bank assets, a broad class of contracts that promise redemption at par on deposits at $t = 1$ would lead to a coordination problem amongst depositors and thus to similar qualitative results. The creditor bank's payoffs are complicated by the fact that they depend on the condition of the debtor bank. If the debtor bank were to become insolvent in period 1 (i.e. condition (1) holds), then the creditor bank receives no money from the debtor bank in period 2, and has to divide up a smaller pool of proceeds amongst its residual claimants. However, regardless of the condition of the debtor bank, the creditor bank may itself be run out of business. Let n_c denote the proportion of the patient depositors of the creditor bank who choose to withdraw in period 1. Observe that if

$$n_c \geq \frac{(1-w)r + (x-D)}{1-(w-x)} \quad (2)$$

the creditor bank shall become insolvent. It is thus possible that the creditor bank shall become insolvent while the debtor bank remains solvent. In the simplest possible interpretation of bankruptcy laws, we assume that in this event the proceeds from the debtor bank will be divided equally amongst all the depositors of the creditor bank. The payoffs to depositors of the creditor bank are:

$$u_W(\theta_i, n_c) = \begin{cases} u[1] & \text{if } n_c < \frac{(1-w)r+(x-D)}{1-(w-x)} \\ u[\frac{(w-D)+(1-w)r}{(w-x)+(1-(w-x))n_c} + g(\theta_{-i}, n_d)] & \text{if } n_c \geq \frac{(1-w)r+(x-D)}{1-(w-x)} \end{cases}$$

$$u_R(\theta_i, n_c) = \begin{cases} u[\frac{(x-D)-(1-w+x)n_c+(1-w)R(\theta_i)+g(\theta_{-i}, n_d)}{(1-n_c)(1-w+x)}] & \text{if } n_c < \frac{x-D}{1-w+x} \\ u[\frac{1-w - \frac{D-x+n_c(1-(w-x))}{r}}{(1-n_c)(1-(w-x))}R(\theta_i)+g(\theta_{-i}, n_d)] & \text{if } \frac{x-D}{1-w+x} \leq n_c < \frac{(1-w)r+(x-D)}{1-(w-x)} \\ u[g(\theta_{-i}, n_d)] & \text{if } n_c \geq \frac{(1-w)r+(x-D)}{1-(w-x)} \end{cases}$$

2.5 A Note on the Payoffs

It is worth discussing some assumptions implicit in the payoffs. The payoffs are written under the assumption that in case the debtor bank's depositors choose to run the creditor bank cannot "pre-empt" them by withdrawing its deposit before the run takes place. This is a reasonable assumption, since our bank runs are induced by local news about local asset shocks, and they are short-lived events. For example, rumors of speculation in the New York money market are likely to be known in New York before they are known in the interior of the country. In addition, relaxing this assumption will not change the qualitative nature of the results. All that is necessary for the results below is that the creditor bank loses *part* of its deposit in case of a run on the debtor bank. This would always be the case except in the unrealistic scenario where the creditor bank knew *before* local depositors at the debtor bank of local news regarding asset returns.

We have also assumed above that payoffs to the debtor banks depositors to be independent of events in the creditor bank. This simplifies the analysis substantially. The content of this assumption is two-fold. First, it means that the debtor bank always pays its interbank claim in period 2 (if it survives) regardless of whether the creditor bank survives or not. The failure of the creditor bank implies, *by definition*, that there are residual claimants and hence it is reasonable for them to be paid from residual assets available at $t = 2$.⁵ The second implication of the assumption has substance. We are implicitly abstracting from a different interlinkage: if the creditor bank experiences a run at $t = 1$, it may be possible for the creditor bank to liquidate its interbank holdings in the debtor bank in period 1. This could generate contagion from the creditor bank to the debtor bank, in addition to the form of contagion (from debtors to creditors) identified in the remainder of the paper.⁶ We argue in Section 4 below, and illustrate in Appendix B, that such contagion from creditors to debtors is likely to be of a much smaller magnitude than contagion in the opposite direction. Thus, our assumption, while restrictive, enables us to isolate the more relevant form of contagion.

2.6 Notation

We label the entire game Γ . For $i = c, d$, we label the realization of Γ in which the depositors of bank i are chosen to act first by Γ_i . Within each Γ_i there are two stage games. We denote the first stage games by $\Gamma_{i,1}$. We denote the second-stage game by $\Gamma_{j,k}$, where

⁵If debtor banks did not pay depositors of failed creditor banks, then creditor failure would benefit debtors. This would not conflict with the debtor-to-creditor contagion results below. The banks' preferences for interbank deposits identified in Section 5 are also qualitatively unchanged: debtor banks find them all the more useful, while to creditor banks they are riskier.

⁶I am grateful to Elu von Thadden for pointing this out.

$j = c, d$ denotes the bank whose depositors make their choices second, and $k = S, F$ denotes whether the bank in the first stage game survived or failed. For example, in the game Γ_d , the stage game $\Gamma_{d,1}$ involves depositors at the debtor bank. This is followed by the stage game $\Gamma_{c,S}$, or $\Gamma_{c,F}$ amongst depositors of the creditor bank, depending on whether the debtor bank survived or failed at stage one.

The structure of the game is common knowledge amongst participants. We look for symmetric Bayesian Nash equilibria of this game.

3 Equilibrium

In order to cleanly characterize the equilibrium set we make the following assumptions:

Assumption 1 (Lower Dominance) *For each depositor of each bank, in each stage game $\Gamma_{i,j}$ for $i \in \{c, d\}$, $j \in \{1, 2\}$, if $\theta_{i,j} < \underline{\theta}$ where $\underline{\theta} \geq L + 2\epsilon$ it is strictly dominant to withdraw.*

In other words, if depositors knew that the bank's returns were going to be sufficiently close to its lowest possible level, it is strictly dominant to withdraw. This is an extremely weak assumption, and emerges essentially endogenously from the payoffs of the game. In addition, since we are principally concerned with the case where $\epsilon \rightarrow 0$, the region $[L, \underline{\theta}]$, which we shall term the *lower dominance region*, can have vanishing measure.

Assumption 2 (Upper Dominance) *For each depositor of each bank, in each stage game $\Gamma_{i,j}$ for $i \in \{c, d\}$, $j \in \{1, 2\}$, if $\theta_{i,j} > \bar{\theta}$, where $\bar{\theta} \leq U - 2\epsilon$ it is strictly dominant to remain.*

In other words, if depositors knew that the bank's returns were going to be sufficiently close to its highest possible level, it is strictly dominant to remain. Again, this region, $[\bar{\theta}, U]$, which we call the *upper dominance region*, can also be as small as we like for $\epsilon \rightarrow 0$. This is also a weak assumption. Though we do not explicitly model it, we can support it by a number of explanations. For example, we could assume that for very high θ , the risky asset in each region pays a premium over cash even in period 1, i.e., the early liquidation payoff r is a function of θ : $r(\theta) < 1$ for $\theta < \bar{\theta}$ but $r(\theta) > 1$ for $\theta \geq \bar{\theta}$. It is also worth emphasizing that the equilibria computed below will exist even without this assumption. However, in that case, there may also be other equilibria.

We wish to also ensure that outside the dominance regions the game does not become trivial. Thus, we insist that even in the high liquidity demand region, there are some patient depositors ($w + x < 1$), and in the complementary region there are some impatient ones ($w - x > 0$). Further, we wish to guarantee that for $\theta \in (\underline{\theta}, \bar{\theta})$, it is possible for banks to fail due to actions by depositors. Inspection of the payoffs above shows that this can be

achieved by imposing the condition that $x < (1 - r)(1 - w)$. Finally, we wish these three conditions to hold simultaneously, which is guaranteed by:

Assumption 3 $x < \min[w, (1 - r)(1 - w)]$

We begin our equilibrium analysis by establishing some intuition for the benchmark case with complete information. With complete information, the signals received by agents are identical to the actual fundamentals. If θ is in the dominance regions, the game has a trivial outcome. However, for $\theta \in (\underline{\theta}, \bar{\theta})$, given Assumption 3, the analysis is more complex. How depositors behave in these states depends crucially on what they believe others are going to do. For any given θ in this intermediate region, if patient depositors believe that other patient depositors will leave their deposits in the bank, then it is optimal for them to do the same. On the other hand, if they believe that a sufficient number of other patient depositors will withdraw prematurely, then it is optimal for them to withdraw too. Thus, a given level of intermediate returns observed by a patient depositor can lead to two distinct “self-fulfilling” outcomes, corresponding to the survival and failure of banks. Such multiplicity was formally modelled for banks in the seminal paper of Diamond and Dybvig (1983).

It is clear from this discussion that outcomes in banking models depend not only on agents beliefs about fundamentals but also on their beliefs about other agents’ actions (and thus, in turn, their beliefs). With complete information, agents have common knowledge of the fundamentals of their bank and complete certainty regarding the actions of other agents in equilibrium. Such strategic certainty enables perfect coordination of actions and beliefs by agents, thus removing the link between fundamentals and outcomes and resulting in multiplicity. The presence of asymmetric information introduces strategic uncertainty and thus reduces the ability of agents to precisely coordinate on arbitrary actions and beliefs. In addition, the existence of the dominance regions also imposes additional structure over beliefs. In the presence of dominance regions the actions of agents with very extreme signals are deterministic, which influences the beliefs of agents with similar but less extreme signals, and in turn, therefore, the beliefs of all agents.

These ideas were first formalized in the context of coordination games by Carlsson and van Damme (1993), and have subsequently been widely applied and generalized (see Morris and Shin (2003) for a survey). Our work belongs to this literature, with one important difference. These papers consider games of strategic complementarities: the excess payoff to taking some action must always be increasing in the number of other agents taking the same action. In our model, when a bank survives, the actions of agents are indeed strategic complements. However, if a bank fails, actions are strategic substitutes. This is because the payoff to running bank diminishes in the event of failure in the number

of other agents who also run the bank. There is a finite pool of resources to be divided up amongst those who choose to run. Nevertheless, we are able to show that in our model there is a unique Bayesian Nash equilibrium. This equilibrium is characterized by *monotone strategies*: strategies under which an agent chooses to remain in the bank if and only if their private information $\theta_{i,j}$ is above some threshold $\theta_{i,j}^*$. Thus each monotone strategy is characterized by a threshold. Symmetric equilibria in such strategies are called *monotone equilibria* or *threshold equilibria*. The stage game amongst the depositors of each bank in the dynamic game is characterized by one such equilibrium threshold, which in turn implies a unique set of thresholds for the dynamic game.

Proposition 1 *There is a unique equilibrium in Γ . In Γ_c it is characterized by the triple $(\theta_{c,1}^*, \theta_{d,S}^*, \theta_{d,F}^*)$. In Γ_d , it is characterized by the triple $(\theta_{d,1}^*, \theta_{c,F}^*, \theta_{c,S}^*)$, where $\theta_{d,S}^* = \theta_{d,F}^* = \theta_{d,1}^*$.*

Here for $\theta_{i,j}^*$ is the threshold for game $\Gamma_{i,j}$ according to the notation introduced above. We now sketch the argument and provide intuition for this result. It is simplest to begin by showing the existence of equilibria in monotone strategies in the stage static games. We illustrate this proof for only one of the static coordination games: the coordination game of the debtor bank's patient depositors. The proofs for the other games are similar.

For the purposes of this proof, denote by θ , the underlying fundamentals of the bank concerned, and by θ_i the signal received by agent i . Upon receiving signal θ_i , the agent has to decide whether to remain or withdraw. The quantity she is interested in is the expected payoff difference between withdrawing and remaining. Suppose all other agents were following threshold strategies with threshold θ^* . Conditional upon receiving signal θ_i , the agent knows that fundamentals lie between $\theta_i - \epsilon$ and $\theta_i + \epsilon$, and has uniform beliefs over this interval. For any θ , therefore, the agent believes that a proportion

$$n(\theta, \theta^*) = \begin{cases} 1 & \text{if } \theta \leq \theta^* - \epsilon \\ \frac{1}{2} + \frac{\theta^* - \theta}{2\epsilon} & \text{if } \theta^* - \epsilon < \theta < \theta^* + \epsilon \\ 0 & \text{if } \theta \geq \theta^* + \epsilon \end{cases}$$

of agents will withdraw from the bank. For a particular (θ, θ^*) , the payoff premium to remaining is given by:

$$\pi(\theta, n) = \begin{cases} u[0] - u\left[\frac{w+D+(1-w)r}{w+x+(1-w-x)n}\right] & \text{if } \frac{(1-w)r+(D-x)}{1-w-x} \leq n \leq 1 \\ u[0] - u[1] & \text{if } \frac{(1-w)r+(D-x)-rD}{1-w-x} \leq n \leq \frac{(1-w)r+(D-x)}{1-w-x} \\ u\left[\frac{1-w-\frac{(1-(w+x)n+(x-D)-D)}{r}}{(1-w-x)(1-n)}R(\theta_i)\right] - u[1] & \text{if } 0 \leq n \leq \frac{(1-w)r+(D-x)-rD}{1-w-x} \end{cases}$$

Thus, the quantity of interest to the agent is

$$\Pi(\theta_i, \theta^*) = \int_{\theta_i - \epsilon}^{\theta_i + \epsilon} \pi(\theta, n(\theta, \theta^*)) d\theta$$

θ^* is a monotone equilibrium if the following hold:

1. $\Pi(\theta^*, \theta^*) = 0$
2. $\Pi(\theta_i, \theta^*) > 0$ if $\theta_i > \theta^*$
3. $\Pi(\theta_i, \theta^*) < 0$ if $\theta_i < \theta^*$

Observe that the existence of the upper and lower dominance regions implies that $\Pi(\theta^*)$ is negative for sufficiently low θ^* and positive for sufficiently high θ^* . Thus, it must cross the θ^* axis somewhere. This establishes (1) above.

To prove (2) and (3) observe that changing θ_i , holding θ^* constant only changes the bounds of integration in $\Pi(\cdot)$. In particular, notice that $\pi(\theta, n) < 0$ for $\theta \leq \theta^* - \epsilon$ and $\pi(\theta, n) > 0$ for $\theta \geq \theta^* + \epsilon$. Since $\Pi(\theta^*, \theta^*) = 0$, the positive and negative parts of the integral exactly offset each other. Increasing θ_i above θ^* increases the positive part of the integral and reduces the negative part, and thus makes $\Pi(\cdot)$ strictly positive. By the same token, reducing θ_i below θ^* makes $\Pi(\cdot)$ strictly negative. Thus, we have established (2) and (3), and this completes the argument for existence.

Given existence, it is not difficult to show, as we do in the appendix, that $\Pi(\theta^*, \theta^*)$ is monotone in θ^* , and thus there is exactly one equilibrium in monotone strategies. This result holds quite generally. Morris and Shin (2003) show that in the absence of strategic complementarities as long as the payoffs satisfy a single crossing property and the information structure satisfies a monotone likelihood ratio property, there exists a unique equilibrium in monotone strategies. With uniform noise, we can show an additional result: there are no other equilibria in more complex non-monotone strategies. The proof technique used to show this result builds on the work of Goldstein and Pauzner (2000a), extending their arguments to our more complex payoffs. It is given in the appendix.

We have thus argued that in each of the stage static coordination games, there is a unique Bayesian Nash equilibrium. What does this imply about equilibria in the dynamic game? The debtor bank depositors' payoff are unaffected by the actions and payoffs of other agents, and thus the debtor bank's depositors' switching threshold is uniquely determined without reference to the actions of depositors of the creditor bank. Therefore $\theta_{d,S}^* = \theta_{d,F}^* = \theta_{d,1}^*$. It is easy to see that corresponding to each possible threshold for the debtor bank, the switching thresholds for the creditor bank's depositors are uniquely determined. Thus, there is a unique equilibrium in the dynamic game.

We conclude this section with an interpretation of this result. For simplicity, consider the case where the bank's fundamentals are almost public, i.e. $\epsilon \rightarrow 0$. In this case, all agents receive essentially the same signal. Let θ^* be the threshold corresponding to the monotone strategies used by the agents in the unique equilibrium for this bank's depositors' game.

If the bank's fundamentals are low, i.e., $\theta < \theta^*$, then essentially all depositors will receive signals below their threshold, and will choose to withdraw their deposits. This, will imply that the bank will fail. In contrast, if $\theta > \theta^*$, no depositors will withdraw, and the bank will survive. Thus, we can reinterpret the switching thresholds of a bank's depositors as the failure thresholds of the banks themselves. In this model, therefore, banks fail due to the (correct) anticipation of adverse returns to their portfolios. Since regional fundamentals are independent, bank failure is therefore driven by *local* shocks to assets. However, we shall see that even such local shocks to bank assets can, via interbank deposits, have larger consequences.

4 Contagion

Contagion emerges as a natural equilibrium property of this game. In the context of bank runs, the most natural concept of contagion is as follows: Consider any two banks within a banking system, i and j . Both banks i and j have some probability of failure independent of what happens in other banks. Thus, even if bank i does not fail, bank j may fail for some realized level of adverse information about it. However, if bank i fails, this may create an adverse effect on bank j . Now, bank j may fail for a *larger* range of information about itself. Thus, we say that the failure of bank i has a *contagious effect* on bank j , if, conditional on the failure of bank i , bank j fails with higher probability than it would have had bank i not failed. Formally, we can define this as follows:

Definition 1 (Contagion) Consider a pair of banks i and j , each with asset returns indexed by θ_i and θ_j . Let $\theta_{j,F}^*$ and $\theta_{j,S}^*$ denote the failure threshold of bank j conditional on the failure and survival of bank i respectively. We say that the failure of bank i contagiously affects bank j if the region of fundamentals $[\theta_{j,S}^*, \theta_{j,F}^*]$ has positive measure.

Having thus defined contagion, we are ready to state a central result of this paper.

Proposition 2 (Contagion) *In the game where depositors of the debtor bank act first, the failure of the debtor bank contagiously affects the creditor bank, i.e., there exists a region of fundamentals $[\theta_{c,S}^*, \theta_{c,F}^*]$ in which the creditor bank fails if and only if the debtor bank fails.*

The proof is in the appendix.

In other words, in the unique equilibrium of our model the failure of debtor institutions adversely affects the prospects of creditor institutions, and thus, *ceteris paribus*, makes it likelier that the creditor shall fail. This is because the failure of the debtor bank reduces the assets of the creditor bank. Rational depositors, knowing this, will be more likely to run on the creditor bank in the event of the failure of the debtor bank. Conditional upon

the failure of a bank, this result also characterizes which other banks, *ceteris paribus*, are likelier to face runs.

It is apparent that under the particular payoff structure imposed here, there is no contagion from creditors to debtors in this model. In a more general model, the failure of a creditor institution could contagiously affect the debtor, if in the process of facing a run from its own depositors, the creditor bank withdraws its deposit from the debtor bank in period 1. However, it is easy to see that such contagion would be of smaller order than the debtor to creditor contagion identified in the result above. The reason is straightforward. The incentives to run a bank for any given depositor is decreasing in the extent of available resources in period 2. The complete failure of the debtor bank in period 1 causes the creditor bank's depositors to lose an amount $DR(\theta)$ at $t = 2$. Consider instead the losses to the depositors of the debtor bank if the creditor bank failed due to a run, in the process prematurely liquidating its deposit of size D in the debtor bank. If the creditor bank had been unaffected by a run, the debtor bank would have owed the creditor bank $DR(\theta)$ at $t = 2$. However, in case of the premature withdrawal of funds by the creditor bank, the debtor no longer has to pay an interbank claim in period 2, but the early withdrawal of D in period 1 (which has to be funded by premature liquidation of the illiquid asset) costs its depositors total resources of $\frac{D}{r}R(\theta)$ at $t = 2$. Thus, losses due to the failure of the creditor to the depositors of the debtor are: $(\frac{D}{r} - D)R(\theta)$ which is strictly smaller than the losses to depositors of the creditor bank due to the failure of the debtor when $r > \frac{1}{2}$. In appendix B below, we compute the payoffs of the debtor bank condition on the failure of the creditor bank in the case where the latter may prematurely withdraw its deposits from the former. We compare the contagion thus caused to the contagion identified above. While a full analytical comparison proves intractable, numerical computations show that for reasonable values of the early liquidation payoff r , debtor-to-creditor contagion is substantially larger than contagion in the reverse direction.

Finally, we demonstrate a natural comparative statics result.

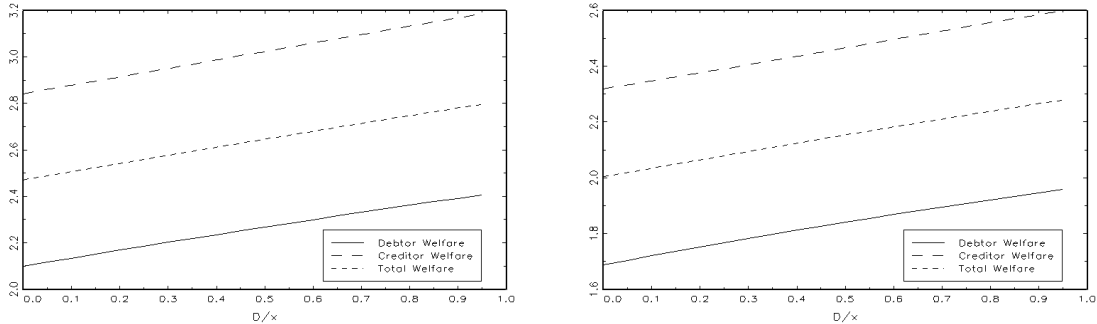
Proposition 3 *The extent of contagion $(\theta_{c,F}^* - \theta_{c,S}^*)$ is increasing in the size of interbank deposit holdings D .*

The proof is in the appendix. The interpretation is straightforward. Contagion flows from debtors to creditors through the channels of interbank deposits. The larger the interbank deposits, the larger the “pipe” through which the contagious effect can flow. This is a testable implication of the model.

5 Should banks hold interbank deposits?

We have shown above that when banks cross-hold deposits to hedge against regional liquidity shocks, the failure of one bank may contagiously affect the other. Thus, in deciding the amount of interbank deposit holdings, banks trade off the benefit of insuring liquidity shocks against the cost of exposing themselves to the risk of contagion. We have demonstrated that for a given set of parameters of our stylized banking system (w, x, r, U, L) , for each choice of $D \in [0, x]$, there is a unique equilibrium with an associated level of social welfare. This makes it possible to determine the optimal interbank deposit amount by maximizing ex ante social welfare.

Upon inspection of the defining equations for the failure thresholds of the banks it is clear that θ_d^* and $\theta_{c,F}^*$ are defined locally with no reference to L or U , the bounds of fundamentals. $\theta_{c,1}^*$ and $\theta_{c,S}^*$ are, in turn, decreasing in U , holding L fixed, since for a given θ_d^* , the higher is U , the higher is the interbank payment received by patient depositors of the creditor bank if they wait until $t = 2$; thus the lower are the incentives to run. Clearly, then, as U gets larger (holding L fixed) the banking system becomes stable, and the relative probability of failure in any bank diminishes. Intuition would suggest that it should then be optimal for banks to fully insure ex ante against liquidity shocks. Figure 1 presents an example to support this intuition.



$$w = 0.3, x = \frac{w}{2}$$

$$w = 0.5, x = \frac{w}{4}$$

Figure 1: Bank Runs Rare: $U = 30$ $L = 0$

We consider risk neutral depositors when bank runs are rare: $L = 0, U = 30; r =$

0.7, $R(\theta) = \sqrt{\theta}$, in the limiting case where noise vanishes (ϵ is set to 10^{-3}). The left panel considers the case in which $w = 0.3, x = \frac{w}{2}$; The right panel sets $w = 0.5, x = \frac{w}{4}$. On the horizontal axis we plot the ratio of interbank deposits (D) to the size of regional liquidity shocks (x). The vertical axis shows the ex ante welfare corresponding to the chosen level of $\frac{D}{x}$. The top and bottom locii represent the ex ante welfare of banks under the hypothetical assumption that they know whether they are going to be interim debtors or creditors (i.e., receive high or low idiosyncratic regional shocks in period 1). Since the two regions receive idiosyncratic liquidity shocks with equal probability in the model, the ex ante welfare locus is simply the arithmetic average of these two locii.

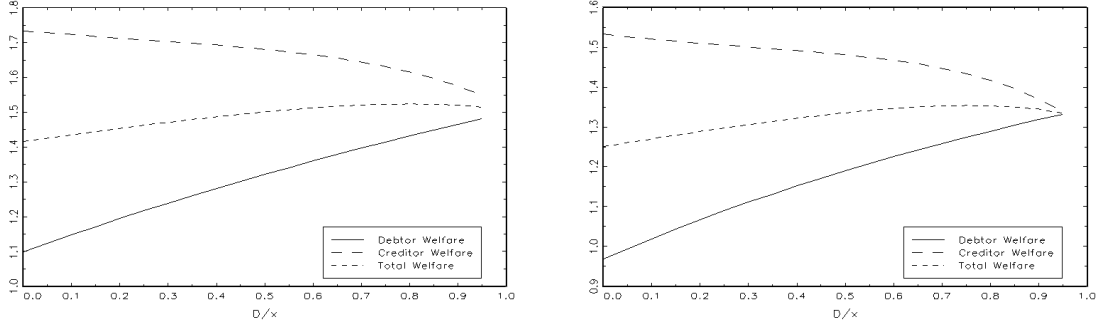
Interbank holdings are always beneficial for the debtor bank. They protect the debtor against liquidity shocks but do not come at the cost of contagion risk. Thus, welfare for the debtor bank is always increasing in D . In the presence of contagion risk, this may not be the case for the creditor bank. It must trade off the benefits of holding interbank deposits (higher returns compared to cash when repaid) against the costs (losses due to debtor failure). However, when bank runs are very rare, as in this example, contagion risk is negligible, and thus creditor bank welfare is also increasing in D . Since both welfare locii are increasing, ex ante welfare is maximized at maximal interbank deposit holdings.

Our framework allows us to illustrate a further result: that banks may sometimes find it optimal *not* to insure fully against regional liquidity shocks. This can occur because while interbank deposits are always better for the interim debtor bank, they are not unambiguously beneficial for the creditor bank, exposing it to contagion risk. When bank runs are less uncommon, this potential risk can affect choices. Since banks assign positive probability to being the interim creditor, they may find it optimal to insure incompletely against regional liquidity variations when bank failures are likely. Figure 2 below illustrates this point. By setting $U = 10$, we now substantially increase the relative likelihood of bank runs, and recompute ex ante social welfare. Here, as before, debtor welfare is increasing in D . However, creditor welfare is decreasing in D . The aggregate effect leads to a non-monotone ex ante welfare function. Thus, banks will find it optimal not to insure completely against liquidity demand shocks. In summary:

Remark 1 When the banking system is relatively unstable, it is not optimal for banks to insure completely against regional liquidity shocks using interbank deposit holdings.

It is thus clear that in relatively unstable banking systems interbank deposits alone cannot eliminate liquidity risk due to regional variations in cash demand. Even if banks had the ability to insure completely against such shocks via interbank deposits, they would choose not to do so.

The preceding discussion suggests that stable and unstable banking systems will have



$$w = 0.3, x = \frac{w}{2}$$

$$w = 0.5, x = \frac{w}{4}$$

Figure 2: Bank Runs Likely: $U = 10$ $L = 0$

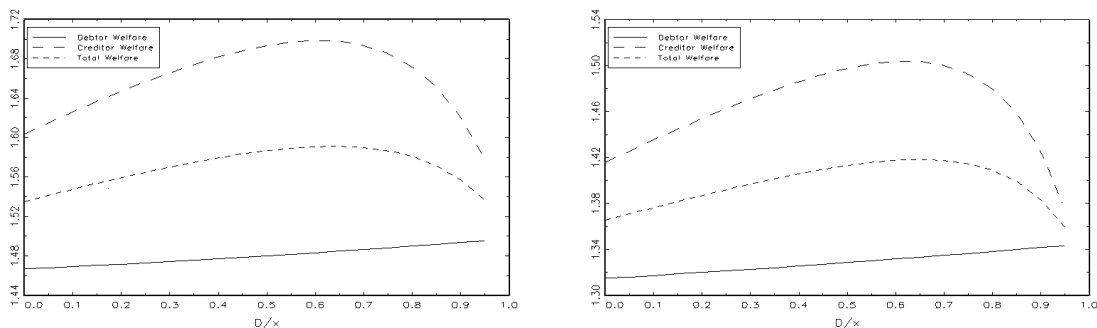
different levels of optimal interconnectedness. In deriving this conclusion we have up to now abstracted from the issue of portfolio choice for banks: corresponding to different levels of interbank deposit holdings, banks may choose to hold excess liquid reserves (bigger than w) to protect themselves against liquidity risk. It is not difficult to see, as we will argue below, that incorporating such “idle reserve” holdings will reinforce the conclusions above. The reason is that interbank deposits and idle reserves are substitutes: they both insure the bank against excess liquidity demand. They are both costly. Idle reserves come with the direct cost of lowered returns to depositors. Interbank deposits come with the indirect cost of potential contagion.

When bank runs are extremely rare (as in the scenario depicted in Figure 1) interbank deposits are clearly preferable to idle reserves. We can numerically solve the model to account for endogenous choice of idle reserves. To be precise, for each level of interbank deposits, we can solve for the ex ante socially optimal level of idle reserves (l), for $D+l \leq x$, thus allowing banks to complement low interbank deposit holdings by high idle reserves. We then compute social welfare for such optimal (D, l) pairs.⁷ Our computations confirm this intuition: even when idle reserves are chosen endogenously, in cases where bank runs are sufficiently rare, banks will optimally use full interbank deposit holdings to insure against

⁷The model is intractable for the unconstrained case where we allow $D+l > x$. However, since there will always be a tradeoff between D and l , we believe that the conclusions will not be qualitatively different.

liquidity shocks and avoid holding idle reserves. Allowing for idle reserves, therefore, does not change Figure 1. Such reserves are most costly precisely in stable banking systems characterized by high returns to bank assets.

However, when bank runs are more common, the situation changes. Consider, for example, the scenario depicted in Figure 2. In this case, U is low and thus interbank deposits are more risky ex ante. At the same time, since returns from illiquid investments are lower, idle reserves are less costly. Thus, intuition suggests that banks may optimally complement interbank deposits by idle reserves. Simulations of the model for the parameters of Figure 2 where idle reserves are selected optimally ex ante confirm this intuition. These results are presented in Figure 3. To obtain some intuition for Figure 3 observe that when the creditor bank is able to complement risky interbank deposits by a buffer of idle reserves, it is welfare-enhancing up to a point to exchange deposits. However, very high levels of interbank deposits are still too risky. The optimum is attained at a lower level of interbank deposits than in the case without idle reserves. This means that in unstable banking systems, banks would ex ante not only complement their interbank holdings with idle reserve buffers, but these buffers to some extent also *substitute* for cross-held deposits. Thus, taking the bank's choices of idle reserves into account accentuates the difference between stable and unstable banking systems discussed above.



$$w = 0.3, x = \frac{w}{2}$$

$$w = 0.5, x = \frac{w}{4}$$

Figure 3: Welfare with endogenous reserves in an unstable system: $U = 10$ $L = 0$

In this context, it is worth making an additional point. Our discussion to date suggests

that in relatively unstable banking systems, banks will optimally choose lower levels of interconnectedness and in addition will hold idle reserves to protect themselves. This suggests that though bank failures are more common in unstable systems by definition, contagious effects, while commonplace, will be of moderate magnitude. However, when banks fail (with low probability) in relatively stable banking systems, contagion, though rare, will be severe: institutions will be highly interconnected in equilibrium, and banks will not have taken precautionary measures in the form of idle reserves to protect themselves. To summarize:

Remark 2 Contagion is much less common in stable banking systems than in unstable ones. However, in the unlikely incidence of bank failure within a relatively stable system, contagion is most severe, since banks have optimally chosen higher levels of interconnectedness *ex ante*, and have optimally not held liquidity buffers in the form of idle reserves.

6 Conclusion

The existence of contagion in the real world is a much debated issue, both in the context of banking systems and more generally (see Gorton and Winton (2002) and Dungey et al (2003) for a survey of the literature). In the context of this debate, this paper makes two contributions. First, we argue theoretically that contagion may occur with positive probability in a banking system due to balance sheet connections across institutions. Such contagion does not require any aggregate excess demand for liquidity in the system, and is purely a spread of local asset-side disturbances. Second, we show that the positive probability of such contagion prevents banks from perfectly insuring each other against liquidity risk via the cross-holding of deposits.

We conclude with a few thoughts on the robustness of these results, and on potential extensions. Our model extends naturally to more than two regions. With more than two regions, holding aggregate liquidity constant, there would be some level of negative correlation across regional liquidity demands. This would create, as before, the incentive to insure against regional liquidity demand shocks using interbank deposits, and thus interim debtor-creditor relationships amongst banks. The only substantive difference would be one of algebraic complexity.

Adding aggregate liquidity shocks to our model creates a second source of bank failure without changing the internal structure of interbank deposits and contagion. With large aggregate liquidity shocks, banks may fail simply because there is just not enough money in the system to meet all claims in period 1 even without expectations-based runs. We limit our attention to constant aggregate liquidity economies and show that contagion occurs even in such economies.

Our analysis abstracts from informational contagion. We assume that fundamentals in the two regions are independent, and thus eliminate any conclusions that agents in one region can draw about their own bank from the observed failure or survival of a bank in a different region. Introducing correlations amongst assets across the regions of our economy would introduce learning into our model and a second source of contagion, enriching the analysis. Incorporating learning into a model similar to ours, building on recent theoretical work on coordination games with social learning (for example, Dasgupta (2002)), is a natural direction for future research.

Appendix A: Proofs

Unique equilibrium in monotone strategies: Again, we prove this only for the coordination game of the patient depositors of the debtor bank, and extend by symmetry to all other games. By a slight abuse of notation, we write $\Pi(\theta^*) = \Pi(\theta^*, \theta^*)$. We shall show $\Pi(\theta^*)$ is monotone in θ^* . Write $n^d = \frac{(1-w)r+(D-x)}{1-w-x}$. Note that if $n(\theta, \theta^*) < n^d$, then $\theta > \theta^* + \epsilon(1 - 2n^d)$. Thus, we can express $\Pi(\cdot)$, as a sum of integrals over θ , with limits of integration given by functions of θ^* , following the piecewise definition of $\pi(\theta, n)$ above. Since the limits of integration are always linear with slope 1 in θ^* , integrating over constant terms gives us final products that are independent of θ^* . Thus, we can rewrite $\Pi(\cdot)$ as

$$\int_{\theta^* + \epsilon(1-2r)}^{\theta^* + \epsilon} f(\theta, \theta^*) d\theta - \int_{\theta^* - \epsilon}^{\theta^* + \epsilon(1-2n^d)} g(\theta, \theta^*) d\theta + K$$

where K proxies for the terms that do not involve θ^* , and

$$f(\theta, \theta^*) = u\left[\frac{1-w - \frac{(1-(w+x))n_d + (x-D)}{r} - D}{(1-w-x)(1-n_d)} R(\theta)\right]$$

and

$$g(\theta, \theta^*) = u\left[\frac{w + D + (1-w)r}{w + x + (1-w-x)n(\theta, \theta^*)}\right]$$

Holding other parameters constant, we differentiate with respect to θ^* :

$$\frac{d}{d\theta^*} \Pi(\theta^*) = \frac{d}{d\theta^*} \int_{\theta^* + \epsilon(1-2r)}^{\theta^* + \epsilon} f(\theta, \theta^*) d\theta - \frac{d}{d\theta^*} \int_{\theta^* - \epsilon}^{\theta^* + \epsilon(1-2n^d)} g(\theta, \theta^*) d\theta$$

Since the limits of integration, in each case are linear in θ^* , their derivatives are simply unity, and thus differentiating under the integral:

$$\frac{d}{d\theta^*} \int_{\theta^* + \epsilon(1-2r)}^{\theta^* + \epsilon} f(\theta, \theta^*) d\theta = f(\theta^* + \epsilon, \theta^*) - f(\theta^* + \epsilon(1-2r), \theta^*) + \int_{\theta^* + \epsilon(1-2r)}^{\theta^* + \epsilon} \frac{d}{d\theta^*} f(\theta, \theta^*) d\theta$$

We can rewrite this to be:

$$\frac{d}{d\theta^*} \int_{\theta^*+\epsilon(1-2r)}^{\theta^*+\epsilon} f(\theta, \theta^*) d\theta = \int_{\theta^*+\epsilon(1-2r)}^{\theta^*+\epsilon} \frac{d}{d\theta} f(\theta, \theta^*) d\theta + \int_{\theta^*+\epsilon(1-2r)}^{\theta^*+\epsilon} \frac{d}{d\theta^*} f(\theta, \theta^*) d\theta$$

Similarly,

$$\frac{d}{d\theta^*} \int_{\theta^*-\epsilon}^{\theta^*+\epsilon(1-2n^d)} g(\theta, \theta^*) d\theta = \int_{\theta^*-\epsilon}^{\theta^*+\epsilon(1-2n^d)} \frac{d}{d\theta} g(\theta, \theta^*) d\theta + \int_{\theta^*-\epsilon}^{\theta^*+\epsilon(1-2n^d)} \frac{d}{d\theta^*} g(\theta, \theta^*) d\theta$$

Now, we observe that (1) $f(\theta, \theta^*)$ decreases in $n(\theta, \theta^*)$, (2) $g(\theta, \theta^*)$ decreases in $n(\theta, \theta^*)$, (3) $n(\theta, \theta^*)$ increases in θ^* , (4) $n(\theta, \theta^*)$ decreases in θ , (5) $|\frac{dn(\theta, \theta^*)}{d\theta}| = |\frac{dn(\theta, \theta^*)}{d\theta^*}|$, since θ and θ^* enter $n(\theta, \theta^*)$ symmetrically, and (6) $R(\theta)$ increases in θ , but is unaffected by θ^* . (1) and (3) imply that $f(\theta, \theta^*)$ decreases in θ^* . (1) and (4) imply that $f(\theta, \theta^*)$ increases in θ . (1), (3), (4), (5), and (6) imply that $|\frac{df(\theta, \theta^*)}{d\theta}| > |\frac{df(\theta, \theta^*)}{d\theta^*}|$. Thus,

$$\frac{d}{d\theta^*} \int_{\theta^*+\epsilon(1-2r)}^{\theta^*+\epsilon} f(\theta, \theta^*) d\theta > 0$$

Similarly, (2) and (3) imply that $g(\theta, \theta^*)$ decreases in θ^* . (2) and (4) imply that $g(\theta, \theta^*)$ increases in θ . (2), (3), and (5) imply that $|\frac{dg(\theta, \theta^*)}{d\theta}| = |\frac{dg(\theta, \theta^*)}{d\theta^*}|$. Thus,

$$\frac{d}{d\theta^*} \int_{\theta^*-\epsilon}^{\theta^*+\epsilon(1-2n^d)} g(\theta, \theta^*) d\theta = 0$$

In the net, we have just shown that $\Pi(\cdot)$ is strictly increasing in θ^* . Thus, there is only one value of θ^* that solves $\Pi(\theta^*, \theta^*) = 0$. ■

No non-monotone equilibria: We present the proof for the static coordination game for the debtor bank's patient depositors. The proofs for all other static games are similar. We first establish a series of lemmas:

Lemma 1 *Let $n(\theta)$ be any feasible belief about the number of patient depositors who choose to run when the state is θ . Then $\frac{dn(\theta)}{d\theta} \in [-\frac{1}{2\epsilon}, \frac{1}{2\epsilon}]$*

Proof: At state θ , the possible realizations of signals lie in $[\theta - \epsilon, \theta + \epsilon]$. Let $p(\theta_i)$ denote the beliefs of agent i about the mass of patient agents who shall run when she receives signal θ_i . Then, for this agent:

$$n(\theta) = \int_{\theta-\epsilon}^{\theta+\epsilon} p(\theta_i) \frac{1}{2\epsilon} d\theta_i$$

Differentiating relative to θ , we have:

$$\frac{dn(\theta)}{d\theta} = \frac{1}{2\epsilon} [p(\theta + \epsilon) - p(\theta - \epsilon)]$$

Since $p(\cdot) \in [0, 1]$, the result follows. ■

Lemma 2 Assume that $0 < \theta_T - \theta_B \leq 2\epsilon$ and that for all $\theta \in [\theta_B, \theta_T]$ $\hat{\theta}(\theta) < \theta_B$ and $n(\theta) \geq \frac{\theta_T - \theta}{2\epsilon}$.

$$\text{If } \int_{\theta_B}^{\theta_T} \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) d\theta \geq 0, \text{ then } \int_{\theta_B}^{\theta_T} \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) d\theta > \int_{\theta_B}^{\theta_T} \pi(\hat{\theta}(\theta), n(\theta)) d\theta$$

Proof: As we have seen above

$$\pi(\theta, n) = \begin{cases} u[0] - u[\frac{w+D+(1-w)r}{w+x+(1-w-x)n(\theta, \theta^*)}] & \text{if } \frac{(1-w)r+(D-x)}{1-w-x} \leq n \leq 1 \\ u[0] - u[1] & \text{if } \frac{(1-w)r+(D-x)-rD}{1-w-x} \leq n \leq \frac{(1-w)r+(D-x)}{1-w-x} \\ u[\frac{1-w-\frac{(1-(w+x))n_d+(x-D)-D}{r}}{(1-w-x)(1-n_d)} R(\theta_i)] - u[1] & \text{if } 0 \leq n \leq \frac{(1-w)r+(D-x)-rD}{1-w-x} \end{cases}$$

Notice the following:

1. When $\frac{(1-w)r+(D-x)}{1-w-x} \leq n \leq 1$, $\frac{\partial \pi(\theta, n)}{\partial \theta} = 0$, $\frac{\partial \pi(\theta, n)}{\partial n} > 0$. Call this the (strategic) “substitutes” range of $\pi(\cdot, \cdot)$.
2. When $\frac{(1-w)r+(D-x)-rD}{1-w-x} \leq n \leq \frac{(1-w)r+(D-x)}{1-w-x}$, $\frac{\partial \pi(\theta, n)}{\partial \theta} = 0$, $\frac{\partial \pi(\theta, n)}{\partial n} = 0$. Call this the “flat” range of $\pi(\cdot, \cdot)$.
3. When $0 \leq n \leq \frac{(1-w)r+(D-x)-rD}{1-w-x}$, $\frac{\partial \pi(\theta, n)}{\partial \theta} > 0$, $\frac{\partial \pi(\theta, n)}{\partial n} < 0$. Call this the (strategic) “complements” range of $\pi(\cdot, \cdot)$.
4. $\pi(\cdot, \cdot)$ is always negative in the “substitutes” or “flat” ranges. The maximum value it can attain in this range is $u[0] - u[w + D + (1 - w)r] < 0$.

Now, suppose $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) > 0$ for all $\theta \in [\theta_B, \theta_T]$. Then the result follows trivially because $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) \geq \pi(\hat{\theta}(\theta), n(\theta))$ for all θ under these circumstances. To see why, notice that since $n(\theta) \geq \frac{\theta_T - \theta}{2\epsilon}$, $\pi(\hat{\theta}(\theta), n(\theta))$ can either fall in the “complements” range, in which case it is smaller than $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon})$ by (3) above, or it can fall in the “flat” or “supplements” range, in which case it is smaller by (4).

Suppose now that $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) > 0$ for some θ and $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) < 0$ for some other θ in $[\theta_B, \theta_T]$. Since $\frac{\theta_T - \theta}{2\epsilon}$ is monotone in θ , there is exactly one point, call it θ_1 at which $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) = 0$. Let

$$\theta_2 = \inf\{\theta \in [\theta_B, \theta_T] : \pi(\hat{\theta}(\theta), n(\theta)) = 0\}$$

Now we shall show that

$$\int_{\theta_B}^{\theta_1} \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) d\theta \geq \int_{\theta_B}^{\theta_2} \pi(\hat{\theta}(\theta), n(\theta)) d\theta \tag{A1}$$

To establish this, we first prove two claims:

Claim 1

$$\pi(\hat{\theta}(\theta), n(\theta)) < 0 \quad \forall \theta \in [\theta_B, \theta_2]$$

Proof of Claim: Consider $\theta < \min[\theta_1, \theta_2]$. For such θ , $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) < 0$. The various possibilities are:

1. $\frac{\theta_T - \theta}{2\epsilon} > \frac{(1-w)r + (D-x) - rD}{1-w-x}$. This implies that $n(\theta) > \frac{(1-w)r + (D-x) - rD}{1-w-x}$. Thus, $\pi(\hat{\theta}(\theta), n(\theta))$ is in the “flat” or “substitutes” range, and is negative.
2. $\frac{\theta_T - \theta}{2\epsilon} \leq \frac{(1-w)r + (D-x) - rD}{1-w-x}$. Now, either $n(\theta) > \frac{(1-w)r + (D-x) - rD}{1-w-x}$, in which the case the above comment applies, or $n(\theta) \leq \frac{(1-w)r + (D-x) - rD}{1-w-x}$, in which case we are in the “complements” range, and by we know that $\pi(\hat{\theta}(\theta), n(\theta)) \leq \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) < 0$.

Thus, if $\min[\theta_1, \theta_2] = \theta_2$ the claim is proved. If $\min[\theta_1, \theta_2] < \theta_2$, then suppose there exist $\theta \in [\min[\theta_1, \theta_2], \theta_2]$ such that $\pi(\hat{\theta}(\theta), n(\theta)) \geq 0$. But, by continuity, then, we can find a point $\theta_3 < \theta_2$ such that $\pi(\hat{\theta}(\theta_3), n(\theta_3)) = 0$, a contradiction. This completes the proof of the claim.

Claim 2

$$n(\theta_2) < \frac{\theta_T - \theta_1}{2\epsilon}$$

Proof of Claim: At θ_1 ,

$$u\left[\frac{1-w - \frac{(1-(w+x))\frac{\theta_T - \theta}{2\epsilon} + (x-D)}{r} - D}{(1-w-x)(1 - \frac{\theta_T - \theta}{2\epsilon})} R(\theta)\right] = u[1]$$

At θ_2 ,

$$u\left[\frac{1-w - \frac{(1-(w+x))n(\theta) + (x-D)}{r} - D}{(1-w-x)(1 - n(\theta))} R(\hat{\theta}(\theta))\right] = u[1]$$

Since $\hat{\theta}(\theta_2) < \theta_1$, it must be the case that $n(\theta_2)$ is smaller than $\frac{\theta_T - \theta_1}{2\epsilon}$ to compensate. This completes the proof of the claim.

Now we shall use Claims (1) and (2) to demonstrate (A1). Denote $m(\theta) = \frac{\theta_T - \theta}{2\epsilon}$. By a change of variables:

$$\int_{\theta_B}^{\theta_1} \pi(\theta, m(\theta)) d\theta = \int_{m(\theta_1)}^{m(\theta_B)} \pi(\theta(m), m) \left| \frac{\partial \theta}{\partial m} \right| dm$$

$$\int_{\theta_B}^{\theta_2} \pi(\hat{\theta}(\theta), n(\theta)) d\theta = \int_{\min\{n(\theta): \theta \in [\theta_B, \theta_2]\}}^{\max\{n(\theta): \theta \in [\theta_B, \theta_2]\}} \pi(\hat{\theta}(\theta(n)), n) \left| \frac{\partial \theta}{\partial n} \right| dn + \int_{\theta: \frac{\partial \theta}{\partial n} = 0} \pi(\hat{\theta}(\theta(n)), n) d\theta$$

The second integral is smaller, because it is computed over a range that is larger (by Claim 2), because $|\frac{\partial \theta}{\partial n}| \geq |\frac{\partial \theta}{\partial m}|$ (by Lemma 1), and because $\pi(\cdot, \cdot)$ is negative in the range considered (by Claim 1). This establishes (A1).

Since $m(\theta)$ declines faster than $n(\theta)$, and by $n(\theta_2) < m(\theta_1)$, we know that $\theta_2 > \theta_1$. Thus,

$$\int_{\theta_1}^{\theta_T} \pi(\theta, m(\theta))d\theta \geq \int_{\theta_2}^{\theta_T} \pi(\theta, m(\theta))d\theta > \int_{\theta_2}^{\theta_T} \pi(\hat{\theta}(\theta), n(\theta))d\theta \quad (\text{A2})$$

We combine (A1) and (A2) to conclude the proof of the lemma. ■

By the existence of the upper and lower dominance regions, we know that for any feasible beliefs n over the actions of other agents, there exists at least one point θ^* such that $\Pi(\theta^*, n) = 0$. If there is only one such point, it must be true that $\Pi(\theta_i, n) > 0$ for all $\theta_i > \theta^*$, and $\Pi(\theta_i, n) < 0$ for all $\theta_i < \theta^*$. But then, there would be only one equilibrium, and it would be a monotone equilibrium with threshold θ^* . Thus, if we could show that under any feasible beliefs n over the actions of other agents, there can be only one point θ^* such that $\Pi(\theta^*, n) = 0$, then we would be able to establish the non-existence of non-monotone equilibria. We now proceed to do so.

Let $\theta_H = \sup\{\theta_i : \Pi(\theta_i, n) \leq 0\}$. By continuity, $\Pi(\theta_H, n) = 0$, and it is easy to see that $\frac{d\Pi(\theta_H, n(\theta))}{d\theta_B} = \pi(\theta_H + \epsilon, n(\theta_H + \epsilon)) - \pi(\theta_H - \epsilon, n(\theta_H - \epsilon))$. By definition of θ_H , $n(\theta_H + \epsilon) = 0$, and thus $n(\theta_H - \epsilon) \geq n(\theta_H + \epsilon)$. Then, it is easy to see that $\frac{d\Pi(\theta_H, n(\theta))}{d\theta_B} > 0$. There must exist a region immediately to the left of θ_H where $\Pi < 0$. If this is a threshold equilibrium, then this region is $[L, \theta_H]$. To the contrary suppose this region is smaller. Let $\theta_L = \sup\{\theta_i : \theta_i < \theta_H, \Pi(\theta_i, n) \geq 0\}$. By continuity, $\Pi(\theta_L, n) = 0$.

Consider the case when $\theta_H - \theta_L < 2\epsilon$.⁸ Then

$$\theta_L - \epsilon < \theta_L < \theta_H - \epsilon < \theta_L + \epsilon < \theta_H < \theta_H + \epsilon$$

Thus, eliminating the common parts of the two integrals, we can write

$$\Pi(\theta_H, n) - \Pi(\theta_L, n) = \int_{\theta_L + \epsilon}^{\theta_H + \epsilon} \pi(\theta, n(\theta))d\theta - \int_{\theta_L - \epsilon}^{\theta_H - \epsilon} \pi(\theta, n(\theta))d\theta$$

Now by a change of variables $\hat{\theta} = \theta - 2\epsilon$, we can re-write this as:

$$\Pi(\theta_H, n) - \Pi(\theta_L, n) = \int_{\theta_L + \epsilon}^{\theta_H + \epsilon} \pi(\theta, n(\theta))d\theta - \int_{\theta_L + \epsilon}^{\theta_H + \epsilon} \pi(\hat{\theta}, n(\hat{\theta}))d\theta$$

Claim 3 *There are two parts:*

⁸The complementary case has an essentially identical, but simpler, proof, and is omitted.

1. For $\theta \in [\theta_L + \epsilon, \theta_H + \epsilon]$, $n(\theta) = \frac{\theta_H + \epsilon - \theta}{2\epsilon}$.

2. $n(\theta_H - \epsilon) \geq n(\theta_L + \epsilon)$

Proof: For $\theta \in [\theta_L + \epsilon, \infty)$, $n(\theta) = Pr(\theta_i \leq \theta_H | \theta)$. This is because if $\theta \in [\theta_L + \epsilon, \infty)$, $\theta_i \in (\theta_L, \infty)$, and the only θ_i for which $\Pi(\theta_i) < 0$ lie in (θ_L, θ_H) . Thus, in particular, for $\theta \in [\theta_L + \epsilon, \theta_H + \epsilon]$, $n(\theta) = Pr(\theta_i \leq \theta_H | \theta) = \frac{\theta_H - \theta + \epsilon}{2\epsilon}$. This proves the first part of the claim.

Using the above, $n(\theta_L + \epsilon) = \frac{\theta_H + \epsilon - \theta_L - \epsilon}{2\epsilon} = \frac{\theta_H - \theta_L}{2\epsilon}$. For $\theta \in (-\infty, \theta_H - \epsilon]$, by an argument parallel to the above, $n(\theta) \geq Pr(\theta_i > \theta_L | \theta)$. The inequality arises because while $\Pi(\theta_i)$ is definitely negative between θ_L and θ_H , it can also be negative elsewhere. Thus, $n(\theta_H - \epsilon) \geq Pr(\theta_i \geq \theta_L | \theta_H - \epsilon) = \frac{\theta_H - \theta_L}{2\epsilon}$. Therefore, $n(\theta_H - \epsilon) \geq \frac{\theta_H - \theta_L}{2\epsilon} = n(\theta_L + \epsilon)$. This proves the second part of the claim.

Given the above claim,

$$\int_{\theta_L + \epsilon}^{\theta_H + \epsilon} \pi(\theta, n(\theta)) d\theta = \int_{\theta_L + \epsilon}^{\theta_H + \epsilon} \pi\left(\theta, \frac{\theta_H + \epsilon - \theta}{2\epsilon}\right) d\theta = \int_{\theta_B}^{\theta_T} \pi\left(\theta, \frac{\theta_T - \theta}{2\epsilon}\right) d\theta$$

where $\theta_B = \theta_L + \epsilon$, and $\theta_T = \theta_H + \epsilon$. It is easy to see that $\int_{\theta_B}^{\theta_T} \pi\left(\theta, \frac{\theta_T - \theta}{2\epsilon}\right) d\theta \geq 0$. Furthermore, note that when $\theta = \theta_L + \epsilon$, $\hat{\theta} = \theta_H - \epsilon$, so that we know, from the second part of the claim that at the bottom end of the integral, $n(\hat{\theta}) \geq n(\theta)$. Thus now, because $\frac{\theta_T - \theta}{2\epsilon}$ decreases at the fastest feasible rate, we can say, for all $\theta \in [\theta_B, \theta_T]$, $n(\hat{\theta}) \geq \frac{\theta_T - \theta}{2\epsilon}$. Finally, note that $\hat{\theta} < \theta$ throughout the bound of intergration. Thus we can now directly apply Lemma 2 to claim that $\Pi(\theta_H, n) - \Pi(\theta_L, n) > 0$, a contradiction, since under our hypotheses $\Pi(\theta_H, n) = \Pi(\theta_L, n) = 0$. \blacksquare .

Proof of Proposition 2: To prove this result, we begin by writing down the threshold equation for the coordination game amongst depositors at the creditor bank conditional on the failure of the debtor bank. First, we write $n_1^c = \frac{x-D}{1-w+x}$, and $n_2^c = \frac{(1-w)r+x-D}{1-w+x}$. Let $l_1 = 1 - 2n_1^c$, and $l_2 = 1 - 2n_2^c$. Finally, for brevity, we let $m = 1 - w + x$, and suppress the arguments of $n(\theta, \theta^*)$. Then, the threshold equation for patient depositors of the creditor bank conditional upon the failure of the debtor bank can be written as $L_f(\theta^*) = R_f(\theta^*)$ where,

$$L_f(\theta^*) = \int_{\theta^* + \epsilon l_1}^{\theta^* + \epsilon l_2} u\left[\frac{1-w - \frac{x-D+nm}{r}}{(1-n)m} R(\theta)\right] d\theta + \int_{\theta^* + \epsilon l_2}^{\theta^* + \epsilon} u\left[\frac{(x-D) - mn + (1-w)R(\theta)}{(1-n)m}\right] d\theta$$

$$R_f(\theta^*) = \int_{\theta^* - \epsilon}^{\theta^* + \epsilon l_1} (u\left[\frac{w-D + (1-w)r}{w-x+mn}\right] - u[0]) d\theta + K_1$$

where $K_1 = \int_{\theta^* + \epsilon}^{\theta^* + \epsilon} u[1]d\theta$. We know by our previous results that there is a unique $\theta_{c,F}^*$ that solves this equation. Now, we write down the corresponding threshold equation for the depositors of the creditor bank conditional upon the survival of the debtor bank as $L_s(\theta^*) = R_s(\theta^*)$, where

$$L_s(\theta^*) = \int_{\theta^* + \epsilon}^{\theta^* + \epsilon} u\left[\frac{(1-w - \frac{x-D+nm}{r})R(\theta) + g}{(1-n)m}\right]d\theta + \int_{\theta^* + \epsilon}^{\theta^* + \epsilon} u\left[\frac{(x-D) - mn + (1-w)R(\theta) + g}{(1-n)m}\right]d\theta$$

$$R_s(\theta^*) = \int_{\theta^* - \epsilon}^{\theta^* + \epsilon} (u\left[\frac{w - D + (1-w)r}{w - x + mn} + g\right] - u[g])d\theta + K_1$$

where K_1 is as before. Observe that since $g > 0$, $L_s(\theta^*) > L_f(\theta^*)$ for all θ^* . Since $u(\cdot)$ is a concave function, $u(x+y) - u(y) \leq u(x) - u(0)$, for all $x, y > 0$. Thus, $R_s(\theta^*) \leq R_f(\theta^*)$ for all θ^* . In particular, this means that

$$L_s(\theta_{c,F}^*) > R_s(\theta_{c,F}^*)$$

i.e., $\theta_{c,F}^* \neq \theta_{c,S}^*$. Now, observe that by analogy to the proof of unique equilibrium in monotone strategies we know that $L_s(\theta^*)$ is increasing in θ^* , while $R_s(\theta^*)$ is invariant with θ^* . Thus, in order to make the indifference equations hold, we need to reduce θ^* below $\theta_{c,F}^*$, and thus, we have just shown that $\theta_{c,S}^* < \theta_{c,F}^*$. ■

Proof of Proposition 3: We refer to the proof of the previous proposition for notation. Clearly, $\theta_{c,S}^*$ is defined by

$$G_s(\theta^*) = L_s(\theta^*) - R_s(\theta^*) = 0$$

while $\theta_{c,F}^*$ is defined by

$$G_f(\theta^*) = L_f(\theta^*) - R_f(\theta^*) = 0$$

By analogy to the proof of earlier results, we note that G_s and G_f are both strictly increasing in θ^* . Also, note that when $g(\cdot) > 0$, it is strictly increasing in D .

We know that $L_s(\theta^*) > L_f(\theta^*)$ for all θ^* . Since $u' > 0$, the difference between the two is increasing in $g(\cdot)$ which, in turn, is increasing in D . We also know that $R_s(\theta^*) \leq R_f(\theta^*)$ because u is concave, and thus $u(x+y) - u(y) \leq u(x) - u(0)$, for all $x, y > 0$. Define $d(y) = u(x+y) - u(y)$, and note that $d'(y) \leq 0$. Thus, the difference between $R_s(\theta^*)$ and $R_f(\theta^*)$ is also weakly decreasing in g , and therefore in D . Thus,

$$G_s(\theta^*) > G_f(\theta^*) \text{ for all } \theta^*$$

and

$$G_s(\theta^*) - G_f(\theta^*) \text{ is increasing in } D \text{ for all } \theta^*$$

Thus, the difference between the zeros of G_s and G_f is increasing in D . ■

Appendix B: Creditor to Debtor Contagion

We now consider the more general setting in which the creditor bank, when faced with a run in period 1, can prematurely liquidate its holdings in the debtor bank at $t = 1$. Suppose the creditor bank has failed, and has in the process withdrawn its interbank deposit in the debtor. What are payoffs to the depositors of the debtor bank in this case? The demand for cash at the debtor bank in period 1, from its own depositors is $w + x + (1 - w - x)n_d$, where n_d , as before, is the proportion of patient depositors who run. The amount of cash available, after the early withdrawal of D by the creditor bank is $w + (1 - w)r$. Thus, if $n_d \geq \frac{(1-w)r-x}{1-w-x}$, the debtor bank fails. Otherwise, it survives. However, if it survives, it no longer has an interbank liability to pay to the creditor bank or its residual claimants. In order to pay the cash demands in period 1, the debtor bank had to liquidate $\frac{w+x+(1-w-x)n_d-w}{r}$ units of the illiquid asset. The remaining proceeds are available solely to its depositors who choose to leave their money in the bank. Thus, payoffs to the depositors are:

$$\begin{aligned}
 u_W(\theta_i, n_d) &= \begin{cases} u[1] & \text{if } n_d < \frac{(1-w)r-x}{1-(w+x)} \\ u\left[\frac{w+(1-w)r}{(w+x)+(1-(w+x))n_d}\right] & \text{if } n_d \geq \frac{(1-w)r-x}{1-(w+x)} \end{cases} \\
 u_R(\theta_i, n_d) &= \begin{cases} u\left[\frac{1-w-\frac{(1-(w+x))n_d+x}{r}}{(1-w-x)(1-n_d)}R(\theta_i)\right] & \text{if } n_d < \frac{(1-w)r-x}{1-(w+x)} \\ u[0] & \text{if } n_d \geq \frac{(1-w)r-x}{1-(w+x)} \end{cases}
 \end{aligned}$$

Using these payoffs, we can now compute the failure threshold of the debtor bank conditional on the prior failure of the creditor bank ($\theta_{d,F}^*$). In addition, from the analysis in the paper, we are able to compute the failure threshold of the debtor bank conditional on the survival of the creditor bank ($\theta_{d,S}^*$). The proportionate difference between these measures the creditor-to-debtor contagion in this economy. The proportionate difference between $\theta_{c,F}^*$ and $\theta_{c,S}^*$ as defined in the paper measures the debtor-to-creditor contagion in this economy. While an analytical comparison of these two proportionate differences proves to be intractable, we can compute and compare the differences numerically. We present in tables 2 and 3 the computations for the two cases identified in the simulations given in Section 5, for the same assumptions on parameter values ($w = 0.3, x = \frac{w}{2}, L = 0, R(\theta) = \sqrt{\theta}$). The specific parameter values do not affect the qualitative properties of our analysis.

The boldfaced entries in each table correspond to the optimal level of interbank deposit holdings, as shown in Section 5. Inspection of the tables makes it clear that contagion from debtors to creditors is much larger in magnitude than contagion in the reverse direction. It is worth noting that all thresholds, except for $\theta_{c,S}^*$ (the threshold of the creditor bank conditional on the survival of the debtor bank) are the same across the two tables. This is not a coincidence. It is because the other thresholds are defined *locally*, and do not depend

$(\frac{D}{x})$	$\theta_{d,S}^*$	$\theta_{d,F}^*$	C→D Contagion	$\theta_{c,S}^*$	$\theta_{c,F}^*$	D→C Contagion
0.2	6.46	7.14	11%	2.03	2.77	36%
0.4	5.84	7.14	22%	2.01	3.61	79%
0.6	5.26	7.14	36%	2.10	4.77	127%
0.8	4.73	7.14	51%	2.34	6.47	175%
1.0	4.36	7.14	64%	2.68	8.27	209%

Table 2: Bank runs relatively likely: $U = 10$

$(\frac{D}{x})$	$\theta_{d,S}^*$	$\theta_{d,F}^*$	C→D Contagion	$\theta_{c,S}^*$	$\theta_{c,F}^*$	D→C Contagion
0.2	6.46	7.14	11%	1.73	2.77	60%
0.4	5.83	7.14	22%	1.42	3.61	155%
0.6	5.26	7.14	35%	1.19	4.77	302%
0.8	4.73	7.14	51%	1.04	6.47	520%
1.0	4.36	7.14	64%	0.99	8.27	731%

Table 3: Bank runs very unlikely: $U = 30$

on the value of U . The value of U affects $\theta_{c,S}^*$ because conditional on the survival of the debtor bank, the creditor bank receives an interbank payment in period 2 from the debtor bank. The higher is U , the higher is the payment it receives, and thus the lower is its failure threshold, for a given level of D .

References

- Acharya, Viral and Tanju Yorulmazer (2002). “Information Contagion and Inter-Bank Correlation in a Theory of Systemic Risk.” Mimeo. London Business School.
- Allen, Franklin and Douglas Gale (2000). “Financial Contagion.” *Journal of Political Economy*, 108(1), pp. 1-33.
- Calomiris, Charles and Gary Gorton (1991). “The Origins of Banking Panics: Models, Facts, and Bank Regulation.” In *Financial Markets and Financial Crises*, edited by Glenn Hubbard, NBER Project Report, Chicago and London: University of Chicago Press.
- Carlsson, Hans and Eric van Damme (1993). “Global Games and Equilibrium Selection.”

Econometrica, 61(5), pp. 989-1018.

Chen, Yehning (1999). “Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities.” *Journal of Political Economy*, 107(5), pp. 946-968.

Dasgupta, Amil (2002). “Coordination, Learning, and Delay.” London School of Economics, Financial Markets Group Discussion Paper No. 435.

Diamond, Douglas and Philip Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity.” *Journal of Political Economy*, 91(3), pp. 401-419.

Dungey, Mardi, Renée Fry, Brenda González-Hermosillo, and Vance Martin (2003). “Empirical Modelling of Contagion: A Review of Methodologies.” Mimeo. Australian National University.

Freixas, Xavier, Bruno Parigi, and Jean-Charles Rochet (2000). “Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank.” *Journal of Money, Credit, and Banking*, 32(3), pp. 611-638.

Goldstein, Itay and Ady Pauzner (2000a). “Demand Deposit Contracts and the Probability of Bank Runs.” *Journal of Finance*, forthcoming.

Goldstein, Itay and Ady Pauzner (2000b). “Contagion of Self-Fulfilling Financial Crises Due to Diversification of Investment Portfolios.” *Journal of Economic Theory*, forthcoming.

Gorton, Gary and Andrew Winton (2002). “Financial Intermediation.” In *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and René Stulz. Amsterdam: North Holland.

Kiyotaki, Nobuhiro and John Moore (1997). “Credit Chains.” Mimeo. London School of Economics.

Kyle, Albert and Wei Xiong (2001). “Contagion as a Wealth Effect.” *Journal of Finance*, 56(4), pp. 1401-1440.

Lagunoff, Roger and Stacey Schreft (2001). “A Model of Financial Fragility.” *Journal of*

Economic Theory, 99(1), pp. 220-264.

Morris, Stephen and Hyun Shin (2003). “Global Games: Theory and Applications.” In *Advances in Economics and Econometrics*, edited by Mathias Dewatripont, Lars Hansen, and Stephen Turnovsky, Cambridge: Cambridge University Press.

Rochet, Jean-Charles and Jean Tirole (1996). “Interbank Lending and Systemic Risk.” *Journal of Money, Credit, and Banking*, 28(4), pp. 733-762.

Rochet, Jean-Charles and Xavier Vives (2000). “Coordination Failure and the Lender of Last Resort.” Mimeo. Toulouse University.

Sprague, O. M. W. (1910). *History of Crises Under the National Banking System*. Washington, D.C.: Government Printing Office.

Wicker, Elmus (2000). *Banking Panics of the Gilded Age*. New York and Melbourne: Cambridge University Press.