

# Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics\*

Amil Dasgupta<sup>†</sup>

London School of Economics

July 2000; Revised November 2002

## 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 contagion is possible in the unique equilibrium of the economy and characterize exactly when it may exist. At the same time, we identify a direction of flow for contagious effects, which provides a rationale for localized financial panics. Simulations identify the optimal level of interbank deposit holdings in the presence of contagion risk. Our results suggest that when the probability of bank failure is low, maximal levels of interbank holdings are optimal. When cross holding of deposits is complete, we demonstrate that the intensity of contagion is increasing in the size of regionally aggregate liquidity shocks.

**JEL Classification Numbers:** G2, C7. **Keywords:** Bank panics, contagion, equilibrium selection, global games.

---

\*I am grateful to V. V. Chari, Timothy Guinnane, Patrick Kehoe, Jonathan Levin, Stephen Morris, Ady Pauzner, Ben Polak, Andreas Roider, Hyun Shin and Debraj Ray for helpful discussions. I thank seminar participants at London Business School, Yale, and at the Maryland meetings of the Econometric Society for comments. Financial support from the Cowles Foundation and the Centre for Mathematical Studies in Economics and Management Science at Northwestern University is gratefully acknowledged. Some of the computations for this project were carried out using Mathematica 3.0.

<sup>†</sup>Mailing address: Room A210, London School of Economics, Houghton Street, London, WC2A 2AE, England. E-mail: a.dasgupta@lse.ac.uk Webpage: <http://fmglse.ac.uk/~amil>

# 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 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. 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.<sup>1</sup>

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.<sup>2</sup>

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 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 makes it difficult to examine even individual bank failures, which then compounds the difficulty of isolating contagious effects in many bank settings.<sup>3</sup> Using and extending some recent developments in the theory of equilibrium selection in coordination games (see Morris and Shin 2000), 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

---

<sup>1</sup>For papers that emphasize this view, see, for example, Kodres and Pritsker (1998), Chen (1999), or Acharya and Yorulmazer (2002).

<sup>2</sup>Two leading papers that emphasize this view are Kiyotaki and Moore (2001) and Allen and Gale (2000). An interesting recent paper which highlights the possibility of contagion through financial links between stakeholders in different institutions, rather than the institutions themselves, is Goldstein and Pauzner (2000b). A related explanation, based on a wealth effect, is offered by Kyle and Xiong (2001).

<sup>3</sup>For the classic multiple equilibrium model of bank runs, see Diamond and Dybvig (1983).

identify contagion precisely and examine its properties.

The set of financial crises that are best represented by our model are the banking panics of the National Banking System in the United States. We now provide a brief and stylized description of these panics.

## 1.1 National Banking Era Panics

The description presented here selectively synthesises and summarizes the descriptions provided by Sprague (1910), Wicker (2000), and Calomiris and Gorton (1991). 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. As late as 1907, Sprague (1910, p. 223) points out that: “New York still maintained its commanding position as a debtor of national banks.”

There were five banking panics of varying intensity in the National Banking Era prior to the Great Depression. They occurred in 1873, 1884, 1890, 1893, and 1907. With the exception of the crisis of 1893, all of these panics began as localized disturbances in New York and subsequently spread to banks in the interior of the country. What were the sources of these panics? Calomiris and Gorton (1991) argue persuasively that the panics typically began with asset-side shocks. Wicker (2000, p. 1) confirms: “In New York, the banking panics began with an unexpected financial shocks . . . the immediate effect being a loss of depositor confidence manifest by bank runs that were bank-specific and sometimes extending to all savings banks.” This was typically followed by suspension of payments by New York banks, followed by suspensions in banks at various parts of the country.

A particularly good example of such contagious panics was the panic of 1907.<sup>4</sup> In 1907, 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 and National Banks 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 linked to the Copper speculators. While some attempts were made to ease the crisis by private bankers led by J. P. Morgan, an unfortunate delay in reaction by the large New York Clearing House led to a widespread

---

<sup>4</sup>Sprague (1910) points out that the panic of 1907 was preceded by no systemic shocks that might conflate our analysis (p. 216): “For our purposes, therefore, we are fortunate in being provided with a crisis which was preceded by no legislation or monetary conditions unfavorable to sound banking.”

panic, followed by several suspensions and bank closures. 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 (or effective suspensions) through much of the country. In the panic of 1907, therefore, we find a clear example of how a financial panic may spread along the channels of direct capital connections between institutions.

To summarize, some of the stylized features of the National Banking System panics are as follows:

- Panics originated due to 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,<sup>5</sup> panics appeared to diffuse nationally through the correspondent network, from debtor New York banks to creditor banks in the interior.

In what follows, we develop a dynamic Bayesian game theoretic model in which many of these stylized features emerge in equilibrium.

## 1.2 Summary of Model and Results

We consider a 3-period economy with two non-overlapping regions, each with a representative competitive bank. Regional banks have access to a storage technology (cash) and a region-specific risky technology that pays a higher expected return than cash if held to maturity, but pays less than par value if liquidated early. The return on the risky asset is revealed in period three, and is an increasing function of uniformly distributed region-specific underlying fundamentals.

There are two continuums of risk-averse consumers, one in each region, each of whom lives three periods. The consumers receive private liquidity shocks: with some probability they may need to consume in period one. They begin life with their endowments deposited in the representative bank of their region. The aggregate level of liquidity demand in the economy is fixed, but the two regions may experience negatively correlated regional shocks. The two representative banks insure against such regional liquidity shocks by holding inter-

---

<sup>5</sup>Wicker (2000) identifies four channels of transmission from New York to the interior, including diffusion via the correspondent network. However, not all of these factors are independent, and at least two of these may formally reduce to consequences of existing correspondent relationships.

bank deposits. Consumer deposits, as well as interbank deposits take the form of standard demand deposit contracts.

Within period one, regional liquidity shocks are realized first and become public knowledge. Then, nature selects the depositors of one of the banks to receive private signals about their bank's fundamentals and make their choices. The depositors of the other bank observe the net proportion of the depositors of the first bank who withdraw their money. Shortly thereafter, the depositors of the other bank receive private information about their bank's fundamentals, and decide whether to remain or withdraw.<sup>6</sup> The seniority of interbank deposits implies that when the depositors of either bank are called upon to make their choices, there is an interim asymmetry between the two banks: one bank (in the region with high initial liquidity demand) is a net debtor to the other bank.

Under weak assumptions on the distribution of fundamentals, we prove that there is a unique threshold in asset returns below which each bank will fail (Propositions 1, 2, 3, and 4). Bank failure thus depends upon the release of adverse information, and the probability of failure is determined endogenously. In our central result, we show that contagion exists: that is, there are regions of fundamentals in which one bank fails if and only if the other bank fails (Proposition 5). Conditional on the failure of the debtor bank, the creditor bank fails for a wider range of its own fundamentals than if the debtor bank survived. However, the failure of the creditor bank does not affect the probability of failure of the debtor bank. Contagion flows from debtors to creditors, and thus spreads along the channels of interbank deposits in a specific direction. Hence, contagion can be localized and not all institutions become potential targets.

Interbank deposits enable banks to hedge regional liquidity shocks, but expose them to the risk of contagion. We illustrate the conditions under which banks would want to hold significant levels of interbank deposits. Intuition suggests and our simulations confirm that when bank runs are rare, financial institutions will insure fully against regional liquidity demand shocks. However, when bank runs are frequent, only partial insurance will be optimal.

Finally, under the assumption of full liquidity insurance, we present comparative statics results to demonstrate that contagion is increasing in the size of regional liquidity demand

---

<sup>6</sup>There are two natural ways to interpret this non-simultaneity in the model, of which we prefer the second. One can think of this model as a discrete approximation of a continuous time model with generically staggered release of information. One can also think of this as a genuine discrete time setting in which (for reasons we do not model) the depositors of one bank suddenly receive information and choose to act. The depositors of the other bank (in another region) learn about the events in the first bank the following day, and then actively seek out information about their own bank immediately thereafter, and choose whether to withdraw or remain.

shocks (Proposition 6). This is a testable implication of the model.

### 1.3 Related Literature

Our paper is connected with a diverse literature. We apply the equilibrium selection techniques summarized in Morris and Shin (2000). Goldstein and Pauzner (2000a) were the first to apply these techniques to the analysis of bank runs. Our model shares features with theirs.<sup>7</sup> 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. Kiyotaki and Moore (2001) explore the method by which contagion flows through credit chains amongst lenders and entrepreneurs. Their model shares with ours the feature that capital 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.

The paper that comes closest to us 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 contains features of Allen and Gale's framework, there are important differences. Allen and Gale work with perfect information and bank panics occur due to aggregate (random) liquidity shocks on the part of the depositors. Such aggregate liquidity shocks are necessary and sufficient for contagion. Our model features incomplete information. Bank runs occur due to adverse information about asset returns. Regional liquidity shocks are necessary *but not sufficient* for contagion.

The existence of contagion in Allen and Gale (2000) requires incompleteness in the interbank deposit market. This may suggest that complete interbank deposits may eliminate contagion, thus reducing the need for a lender of last resort. Our model suggests, however, that contagion occurs with positive probability even with complete interbank deposits.

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 threshold equilibria. Section 4 contains our central result. The optimal level of interbank deposit holdings is illustrated numerically in Section 5. Section 6 provides closed forms and comparative statics. Section 7 discusses and concludes.

---

<sup>7</sup>Importantly, in both models, payoffs fail to satisfy global strategic complementarities.

## 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  $\lambda$  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 symmetric (representative, competitive) banks which lie in two regions of the economy. Agents begin their lives with their endowments deposited in the bank of their region.<sup>8</sup> There are two classes of assets available to banks: 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 residents also have access to risky illiquid technology  $R_i$ , with returns

---

<sup>8</sup>While we thus assume participation in the banking system, it shall become clear below (See section 5) that this assumption can be justified in equilibrium by choosing free parameters  $L$  and  $U$  to make bank runs rare. As bank runs become sufficiently rare, it shall always be optimal for risk averse depositors to deposit their endowment in banks.

given by:

$$R_i(t) = \begin{cases} 0 < r < 1 & \text{when } t = 1, \\ R(\theta_i) & \text{when } t = 2, \text{ where } \theta_i \text{ is distributed uniform on } [L, U] \end{cases}$$

where  $t$  is the time of liquidation,  $R(\cdot)$  is any increasing function. The parameter  $\theta_i$  indexes some underlying “fundamentals” related to the bank’s assets, which determine the level of the bank’s asset returns. These fundamentals  $\theta_i$  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.

Banks are constrained to offer depositors *demand deposit contracts*.<sup>9</sup> 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.<sup>10</sup> 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.<sup>11</sup> Thus, in this symmetric scheme, banks exchange deposits of size  $D$ ,<sup>12</sup> and distribute their net wealth of 1, putting  $w$  in cash,

---

<sup>9</sup>Hence we are not solving here for the optimal contractual form. Demand deposit contracts are a standard feature of banking systems and we take them as given. For theoretical justification of demand deposit contracts, see Calomiris and Kahn (1991) or Diamond and Dybvig (1983).

<sup>10</sup>We are implicitly assuming that banks have access to only interbank deposits as a tool to insure, and hence can insure only against shocks to their liabilities. While this assumption is not central to our analysis, we assume it for algebraic simplicity as well as to eliminate correlation between the bank’s asset returns. Such correlation would introduce a second channel of contagion, through learning. Our purpose in this paper to explore the extent to which capital connections, *by themselves* can contribute to contagion.

<sup>11</sup>Given the cash holdings of the banks, and given the timing of the model to be explicated below, interbank deposits of size larger than  $x$  will not be desirable to banks. Such deposits would leave banks unable to pay their own early consumers in *any* equilibrium without prematurely liquidating some of their long term assets. In practice, there were often regulatory restrictions on the size of  $D$ .

<sup>12</sup>One can think of this as the banking system entering into an implicit liquidity insurance scheme ex ante – no money physically needs to be exchanged in Period 0



and  $1 - w$  in long term investment projects. The size of interbank deposits,  $D$ , is a *choice variable* for banks. In equilibrium,  $D$  shall depend on the parameters of the model.

These interbank deposits have the feature that in each period they “clear before” claims to individual depositors are paid. In other words, as soon as the state  $A$  or  $B$  is realized at the beginning of period 1, 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. Similarly, in period 2, the banks use any proceeds of liquidated investments to first pay their fellow bankers, and then pay their patient depositors out of the remaining proceeds.<sup>13</sup> 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 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  $(w - x)R(\theta_A)$ . 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 shown below in itemized form:

- Period 0
  - Interbank deposits are initiated.

---

<sup>13</sup>Note that the assumed priority order for clearing is an innocuous assumption: 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 acutally works *against* us.

- 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:

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

Now consider the payoffs to the depositors of the creditor bank. Observe that 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.

Conditional on the failure of the debtor bank (i.e., if condition (1) holds), the payoffs are:

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

Conditional on the survival of the debtor bank the payoffs are:

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

is always at least *some* positive proportion of patient agents. Otherwise the game  $\Gamma$  would be vacuous. Hence, we assume that  $w + x < 1$ , or, in other words

$$x < 1 - w \quad (3)$$

Similarly, in the low liquidity demand state, we insist that there are at least some impatient agents, and hence:

$$x < w \quad (4)$$

Finally, to keep the stage coordination games non-trivial, we require that both the debtor bank and the creditor bank can become insolvent for some ranges of withdrawals by patient agents. In other words, we insist that  $\frac{(1-w)r+(D-x)}{1-w-x} < 1$  for all  $D$ . In particular, since the LHS is maximized at  $D = x$ , we require that  $\frac{(1-w)r}{1-w-x} < 1$ , or,

$$x < (1-r)(1-w) \quad (5)$$

Since we wish all the features implied by the above restrictions to hold in our model, we consider parameters in the region implied by their intersection, i.e.,  $(r \in (0, 1), w > 0, x > 0)$  are chosen such that

$$x < \min[w, (1-r)(1-w)] \quad (6)$$

We are now ready to find equilibria of this model.

### 3 Equilibrium

In each of the static coordination games  $\Gamma_{i,1}$  and  $\Gamma_{i,2}$  for  $i = c, d$ , with common knowledge of  $\theta_i$ , there are (at least) two equilibria. In one of these equilibria, all patient agents remain in the bank (tell the truth) because they expect other patient agents to do so. In the other, they all withdraw money from the bank, because they expect other patient agents to do so. With common knowledge of  $\theta_i$ , patient agents have nothing to condition their beliefs upon, and hence *any* feasible belief about the actions and beliefs of others is permissible. However, with private signals, the situation changes.

In the game with incomplete information, agents are able to condition their beliefs on their private signals, which are known to be correlated with the private signals of others. Since  $R(\theta_i)$  is increasing in  $\theta_i$ , and private signals are positively correlated with fundamentals, high signals convey “good news” to patient agents, and, *ceteris paribus*, make remaining in the bank more attractive. A natural class of strategies that emerges are those in which agents actions are increasing in their beliefs: for good information, agents are more optimistic and more easily persuaded to remain in the bank; the opposite is true for bad

information. Since this is a binary action game, such strategies take a particularly simple form: Agents choose to remain in the bank if their private information  $\theta_{i,j}$  is above some threshold  $\theta_{i,j}^*$ , and choose to withdraw otherwise. We shall call such strategies *monotone strategies*, and equilibria in such strategies *monotone equilibria*. Monotone equilibria are characterized by threshold levels, and hence we shall sometimes refer to these as equilibrium thresholds. In what follows, we first restrict attention to such monotone equilibria, and demonstrate their existence and uniqueness. We then show that there are no other more complicated Bayesian Nash equilibria in these games: the unique monotone equilibrium is also the unique equilibrium. However, first, we need to make two weak assumptions, which we shall refer to as the lower and upper dominance assumptions. We explicate these below.

**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  $\theta_{i,j} = L$ , it is strictly dominant to withdraw.*

In other words, if depositors knew that the bank's returns were going to be at 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. Observe that since at  $\theta = L$  the dominance is strict, there is some (possibly vanishing measure) region in the neighborhood of  $L$  over which dominance holds. Call this region  $[L, \underline{\theta}]$ , the *lower dominance region*.

**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  $\theta_{i,j} = U$ , it is strictly dominant to remain.*

In other words, if depositors knew that the bank's returns were going to be at its highest possible level, it is strictly dominant to remain. This is also a weak assumption. It can be supported by a number of explanations. For example, we could assume that for very high  $\theta$ , the risky asset in each region pays a premium over cash even in period 1. We could also argue that when extremely high level of future returns are guaranteed, the bank becomes a very attractive target for potential purchases by larger, more liquid, banks, making it optimal for patient depositors never to withdraw their money early. The strict dominance inherent in this assumption, implies, as before, that there exists some region  $[\bar{\theta}, U]$ , in the interior of which the dominance result holds. We call this the *upper dominance region*. Note that this region can be vanishingly small.

Given these two assumptions, we are able to characterize the equilibrium outcome of each of these static coordination games.

**Proposition 1 (Monotone Existence)** *For each stage game  $\Gamma_{i,j}$  for  $i \in \{c, d\}$ ,  $j \in \{1, 2\}$ , there exists a threshold level in the fundamentals  $\theta^*(\Gamma_{i,j})$  such that agents who receive*

signals above  $\theta^*(\Gamma_{i,j})$  will remain in the bank, while those that receive signals below it will withdraw.

**Proof:** We demonstrate 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 simpler than but otherwise identical to the proof for this one.

For the purposes of this proof, denote by  $\theta$ , the underlying fundamentals of the bank concerned, and by  $\theta_i$  the signal received by agent  $i$ . Upon receiving signal  $\theta_i$ , 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  $\theta^*$ . Conditional upon receiving signal  $\theta_i$ , the agent knows that fundamentals lie between  $\theta_i - \epsilon$  and  $\theta_i + \epsilon$ , and has uniform beliefs over this interval. For any  $\theta$ , therefore, the agent believes that a proportion

$$n(\theta, \theta^*) = \begin{cases} 1 & \text{if } \theta \leq \theta^* - \epsilon \\ \frac{1}{2} + \frac{\theta^* - \theta}{2\epsilon} & \text{if } \theta^* - \epsilon < \theta < \theta^* + \epsilon \\ 0 & \text{if } \theta \geq \theta^* + \epsilon \end{cases} \quad (7)$$

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

$$\pi(\theta, n) = \begin{cases} u[0] - u[\frac{w+D+(1-w)r}{w+x+(1-w-x)n}] & \text{if } \frac{(1-w)r+(D-x)}{1-w-x} \leq n \leq 1 \\ u[0] - u[1] & \text{if } \frac{(1-w)r+(D-x)-rD}{1-w-x} \leq n \leq \frac{(1-w)r+(D-x)}{1-w-x} \\ u[\frac{1-w-\frac{(1-(w+x))n+(x-D)}{r}-D}{(1-w-x)(1-n)}R(\theta_i)] - u[1] & \text{if } 0 \leq n \leq \frac{(1-w)r+(D-x)-rD}{1-w-x} \end{cases} \quad (8)$$

Thus, the quantity of interest to the agent is

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

$\theta^*$  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  $\theta^*$  and positive for sufficiently high  $\theta^*$ . Thus, it must cross the  $\theta^*$  axis somewhere. This establishes (1) above.

To prove (2) and (3) observe that changing  $\theta_i$ , holding  $\theta^*$  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  $\theta_i$  above  $\theta^*$  increases the positive part of the integral and reduces the negative part, and thus makes  $\Pi(\cdot)$  strictly positive. By the same token, reducing  $\theta_i$  below  $\theta^*$  makes  $\Pi(\cdot)$  strictly negative. Thus, we have established (2) and (3). ■

Having thus shown existence of monotone equilibria, we now demonstrate that they are unique:

**Proposition 2 (Monotone Uniqueness)** *For each stage game  $\Gamma_{i,j}$  for  $i \in \{c, d\}$ ,  $j \in \{1, 2\}$ , there is only one threshold level in the fundamentals  $\theta^*(\Gamma_{i,j})$  such that agents who receive signals above  $\theta^*(\Gamma_{i,j})$  will remain in the bank, while those that receive signals below it will withdraw.*

The proof is in the appendix. Given that there is a unique equilibrium in monotone strategies, a natural question is: are there other equilibria, involving the use of more complex strategies? We demonstrate that there are no such complex equilibria. In games similar to our stage games but with full strategic complementarities the work of Carlsson and van Damme and Morris and Shin have demonstrated that there is a unique strategy profile that survives the iterative deletion of dominated strategies. Our payoffs do not satisfy full strategic complementarities: conditional on the failure of a bank, the actions of individual depositors are strategic *substitutes*. Thus, the strong dominance-solvability result does not apply to our setting. However, by restricting attention to *equilibrium* strategies, we are able to show that behavior shall always be monotone. The proof technique used here builds on the work of Goldstein and Pauzner, extending their arguments to our more complex payoffs. Formally, the result states:

**Proposition 3 (No non-monotone equilibria)** *For each stage game  $\Gamma_{i,j}$  for  $i \in \{c, d\}$ ,  $j \in \{1, 2\}$ , there is no non-monotone equilibrium.*

The proof is in the appendix.

Let us now consider these stage games in sequence. To do this, we introduce some additional notation. There are four possible stage games, as we have discussed above. The depositors of the creditor bank may be chosen to receive information and act first. This generates the stage game  $\Gamma_{c,1}$ . Denote the threshold of this game by  $\theta_{c,1}^*$ . The depositors of creditor bank may also move second. In particular, they may do so after observing that the debtor bank has failed or survived. Denote the threshold of the creditor bank's depositors conditional upon the survival of the debtor bank by  $\theta_{c,S}^*$ . Conditional upon the failure of



the debtor bank, call this threshold  $\theta_{c,F}^*$ . Correspondingly denote by  $\theta_{d,1}^*$ ,  $\theta_{d,S}^*$ , and  $\theta_{d,F}^*$  the respective thresholds for the depositors of the debtor bank.

We have shown above that there is a unique equilibrium in each stage game. Does this mean that there is a unique equilibrium in the dynamic game  $\Gamma$ ? In general, these thresholds may be interlinked by intricate functional relationships. Thus, uniqueness in the component static games does not necessarily imply uniqueness in the dynamic game. However, note that a solvent debtor bank is assumed to always pay its debt to the residual claimants of the creditor bank, regardless of whether the creditor bank remains solvent or not.<sup>14</sup> Thus, the thresholds of the debtor bank's depositors are independent of the outcome in the creditor bank, and, therefore, of the thresholds of the creditor bank's depositors. Thus,  $\theta_{d,1}^*$ ,  $\theta_{d,S}^*$ , and  $\theta_{d,F}^*$  are all uniquely defined in the general dynamic game. In particular, they are equal to each other. Write  $\theta_d^*$  for the unique threshold of the debtor bank. Now, by backward induction,  $\theta_{c,1}^*(\theta_d^*)$  is uniquely defined. Similarly,  $\theta_{c,S}^*(\theta_d^*)$  and  $\theta_{c,F}^*$  are uniquely defined. Thus, we have just argued that the following result holds:

**Proposition 4 (Dynamic Uniqueness)** *There is a unique equilibrium in  $\Gamma$ . In  $\Gamma_c$  it is characterized by the triple  $(\theta_{c,1}^*, \theta_{d,S}^*, \theta_{d,F}^*)$ . In  $\Gamma_d$ , it is characterized by the triple  $(\theta_{d,1}^*, \theta_{c,F}^*, \theta_{c,S}^*)$ .*

One straightforward interpretation of propositions 1, 2, 3, and 4 taken together is that banks fail upon the release of adverse news about them. This matches the commonly made observation that bank failures are positively correlated with the release of bank-specific or region-specific adverse information.

Having established uniqueness of equilibria, it is now of interest to us to explore some of their properties.

## 4 Contagion

Contagion emerges as a natural property of the unique equilibrium thresholds 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*

---

<sup>14</sup>This assumption is not just one of convenience but is also the simplest interpretation of bankruptcy law. The failure of the creditor bank by definition leaves a set of residual claimants to its assets. These assets naturally include expected payments from the debtor bank. The qualitative results of the model do not depend on this assumption.

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  $\theta_i$  and  $\theta_j$ . 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 5 (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. But in the game where depositors of the creditor bank act first,  $\theta_{d,F}^* = \theta_{d,S}^*$ . Thus, the failure of the creditor does not contagiously affect the debtor.*

The proof is in the appendix.

In other words, contagious effects flow in a specific direction, from debtors to creditors. The failure of debtors adversely affects the failure of creditors, 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, shall be more likely to run on the creditor bank in the event of the failure of the debtor bank. On the other hand, the failure of a creditor makes the debtor no better or worse off, since we have assumed that the debtor has to always pay the residual claimants of the failed creditor the amount originally owed.<sup>15</sup> Focussing purely on capital linkages as a channel for contagion, this result suggests why individual bank runs may not necessarily become widespread bank panics. Conditional upon the failure of a bank, this theorem also characterizes which other banks, ceteris paribus, are likelier to face runs. Thus, this result provides one rationale for why panics may be “local” in some specific but unobvious sense.

## 5 Should banks hold interbank deposits?

In our discussion to date, we have not commented on the size of interbank deposit holdings. 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

---

<sup>15</sup>Any other assumption would imply that the debtor actually *benefits* from the failure of the creditor, thus naturally ruling out contagion, leaving our results unchanged.

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 monotone 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. While the model is too complex to derive an analytical formula for such optimal interbank deposit levels in general, it is possible to characterize the solution numerically. We shall show later that under additional assumptions motivated by the numerical solutions presented here we can indeed solve the model in closed form and perform comparative statics.

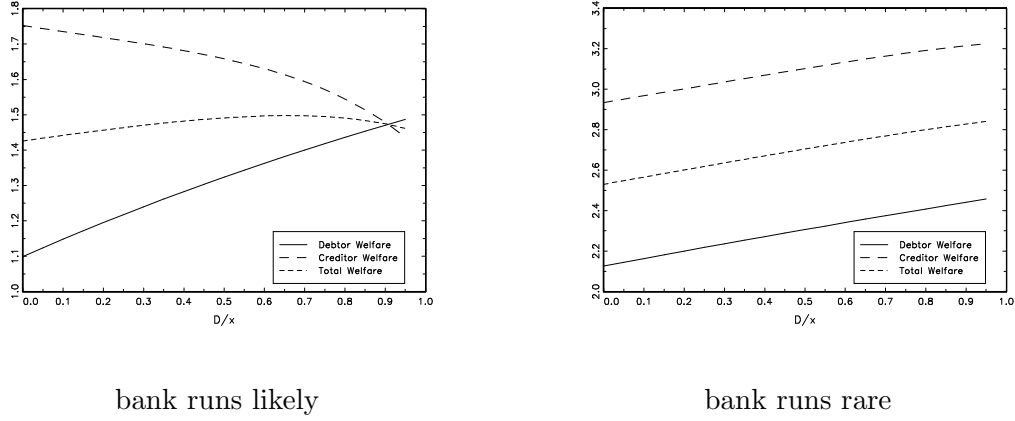
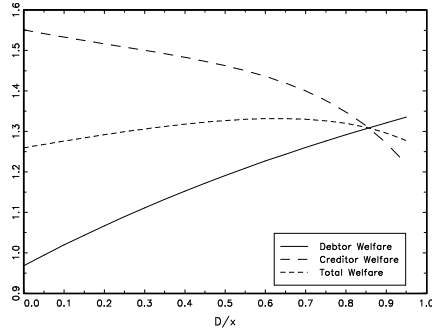


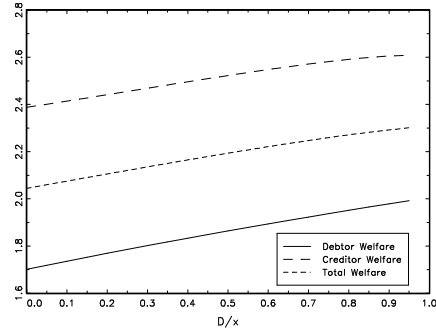
Figure 1: Welfare Comparisons:  $w = 0.3$ ,  $x = \frac{w}{2}$ ,  $r = 0.7$

We present typical results from our simulations in Figures 1 and 2. These illustrate the optimal holding of interbank deposits in a risk-neutral economy. Along the horizontal axis we plot the proportion 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 level of  $\frac{D}{x}$  corresponding to the maximal level of social welfare denotes the optimal holdings of interbank deposits.

In figure 1, we set the constant aggregate liquidity demand ( $w$ ) to be 30%, and the level of regional variation at  $\frac{w}{2}$ . We fix the lower bound of fundamentals ( $\theta$ ) to be  $L = 0$ , and vary the upper bound ( $U$ ) to change the equilibrium probability of bank runs. The left panel in each image corresponds to the case in which bank runs are relatively frequent ( $U = 10$ ), and the right panel to the case in which runs are rare ( $U = 30$ ). The early liquidation return  $r$  is set to 0.7, and  $R(\theta) = \sqrt{\theta}$ . Figure 2 portrays a similar economy in



bank runs likely



bank runs rare

Figure 2: Welfare Comparisons:  $w = 0.5$ ,  $x = \frac{w}{4}$ ,  $r = 0.7$

which average liquidity demand is greater ( $w = 0.5$ ), but the proportional level of regional variation is smaller ( $x = \frac{w}{4}$ ). In all cases, the analysis is carried out “close” to the full information limit. The bound on idiosyncratic noise ( $\epsilon$ ) is set to  $\frac{1}{1000}$ .

The  $x$ -axis in each image shows  $\frac{D}{x}$ , the proportionate size of interbank deposit holdings. The central locus in each figure represents the ex ante social welfare at the corresponding level of  $D$ . From the figures it is apparent that when bank runs are frequent, social welfare is maximized for *intermediate* values of  $D$ , whereas when runs are rare, welfare is maximized for *maximal* values of  $D$ .

Intuition for this property can be derived upon inspection of the other two locii in each of the images. These 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 central welfare locus is simply the arithmetic average of these two locii. It follows from the model that interbank deposit holdings are *always* beneficial for debtors. They are not affected by contagion. Thus interbank holdings insure them against regional liquidity shocks without exposing them to any additional risk. This is evidenced in the figures: the locii representing debtor welfare are always strictly increasing in  $\frac{D}{x}$ . For the bank that is the interim creditor, there is a tangible risk of contagion. However, contagion is a *conditional* event. A necessary condition for contagion is that the debtor bank must actually experience a run. Thus, the

attractiveness of interbank deposits to creditor banks depends crucially on their assessment of the probability of runs on the debtor bank. When this probability is high, a higher level of interbank deposit holdings exposes the creditor to greater contagion risk, thus lowering its ex ante welfare. This can be seen in the left panels of the figures. When the probability that the debtor bank will fail is low, it becomes much more attractive for the creditor bank to hold interbank deposits, which enable it to eliminate idle reserves and increase expected payments to its depositors. Under these circumstances, holding higher interbank deposits increases the welfare of the interim debtor bank. This can be seen in the right panels of the figures.

Ex ante social welfare is given by the average of debtor and creditor welfares. When bank runs are rare, and both debtor and creditor welfare increasing in  $\frac{D}{x}$ , social welfare is maximized at  $D = x$ . Thus, banks should hold a maximal level of interbank deposits. However, when bank runs are more likely, creditor welfare decreases in  $\frac{D}{x}$  while debtor welfare increases. Thus, the social welfare locus is hump-shaped, and there is a strictly interior level of optimal interbank deposit holdings.

## 6 Closed Forms, Comparative Statics

Historically, bank runs have been rare. The simulations presented above suggest that when bank runs are rare, banks will find it optimal to insure against idiosyncratic liquidity shocks by holding a maximal level of interbank deposits. For the remainder of the paper, we shall, therefore, assume that banks fully insure against regional liquidity shocks by setting  $D = x$ .

This assumption enables us to demonstrate further properties of contagion. Under the assumption of complete interbank deposit holdings, we can solve the linear realization of the model (with risk neutral consumers and  $R(\theta) = \theta$ ) in closed form.

Below, we present the closed forms for the equilibrium thresholds for this simplified model. This requires a small intermediate step. The thresholds of depositors of the creditor bank are functions of the interbank payments from the debtor bank. These interbank payments, in turn, are functions of the proportion of premature withdrawals from the creditor bank (which may or may not be observable to the depositors of the creditor bank, depending on whether  $\Gamma_c$  or  $\Gamma_d$  is realized) and the underlying fundamentals of the debtor bank (which are never observable to the depositors of the creditor bank). The depositors of the creditor bank, therefore, have to compute the *expected* interbank payments from the debtor bank. In preparation for writing down the closed forms for equilibrium thresholds, we explicate below the functional form of the expected interbank payments for the game.

## 6.1 Expected Interbank Payments

There are two situations in which payments from the debtor bank must be anticipated by depositors in the creditor bank. In game  $\Gamma_c$  the depositors of the creditor bank must, upon acting first, decide what probability with which the debtor bank will remain solvent, and, if solvent, the expected payment that they shall make. Observe that the depositors of the creditor bank can calculate  $\theta_d^*$ , which is uniquely defined in terms of the parameters. Given  $\theta_d^*$ , they can assign probabilities to ranges of  $n_d(\theta_d, \theta_d^*)$ , and thus compute  $\hat{g} = E(g|\theta_d^*)$ .

In the game  $\Gamma_d$  when the debtor bank survives, the depositors of the creditor bank can observe  $n_d$ . When  $0 < n_d < 1$ , there is a one-to-one relationship between  $n_d$  and  $\theta_d$ , and thus for strictly interior  $n_d$ , depositors at the creditor bank can compute  $g(n_d, \theta_d)$  exactly. However, when  $n_d = 0$ , the depositors of the creditor bank know simply that  $\theta_d > \theta_d^* + \epsilon$ , or, in the limit as noise vanishes,  $\theta_d > \theta_d^*$ . Thus, their expected value of  $\theta_d$  is  $\frac{\theta_d^* + U}{2}$ . Hence, for this game, we can write  $\hat{g}$  as follows:

$$\hat{g}(\theta_d|\theta_d^*, n_d) = \begin{cases} xR(\frac{\theta_d^* + U}{2}) & \text{if } n_d = 0 \\ g(\theta_d, n_d) & \text{if } 1 > n_d > 0 \end{cases} \quad (9)$$

We are now ready to write down the equilibrium thresholds.

## 6.2 Limiting Thresholds

The equilibrium thresholds are particularly tractable in the limit as  $\epsilon \rightarrow 0$ . The limit threshold points of interest are as follows:

- For the coordination game amongst depositors at the debtor bank:

$$\theta_d^* = r \frac{r(1-w) + (r(1-w) + w + x) \ln \left[ \frac{1}{r(1-w) + w + x} \right]}{(1-w-x)(r + (1-r) \ln[1-r])} \quad (10)$$

- For the coordination game amongst depositors at the creditor bank conditional upon the survival of the debtor bank (in the game  $\Gamma_d$ ):

$$\theta_{c,S}^* = r \frac{r(1-w) + \hat{g}(n_d, \theta_d^*) \ln \left[ 1 - \frac{1-w}{1-w+x} r \right] + (r(1-w) + w - x) \ln \left[ \frac{1}{r(1-w) + w - x} \right]}{r(1-w) + (1-r(1-w) - w + x) \ln \left[ 1 - \frac{1-w}{1-w+x} r \right]} \quad (11)$$

where  $\hat{g}(\cdot)$  is defined as in (9).

- For the coordination game amongst depositors at the creditor bank conditional upon

the failure of the debtor bank (in the game  $\Gamma_d$ ):

$$\theta_{c,F}^* = r \frac{r(1-w) + (r(1-w) + w - x) \ln \left[ \frac{1}{r(1-w) + w - x} \right]}{r(1-w) + (1 - r(1-w) - w + x) \ln \left[ 1 - \frac{1-w}{1-w+x} r \right]} \quad (12)$$

Now, it is of interest to us to explore some of the comparative statics of our model by using these closed form expressions for the thresholds. In particular, we examine below how the magnitude of the contagious effect changes as the volatility of the financial system, or, equivalently, the depth of inter-institutional financial linkages increases.<sup>16</sup> It turns out that the degree of the contagious effect increases in the strength of financial linkages between banks. Formally,

**Proposition 6** *The size of contagion increases with the size of the regional liquidity shocks, i.e.,  $ct(r, w, x) = \theta_{c,F}^* - \theta_{c,S}^*$  is increasing in  $x$ .*

The proof is in the appendix.

This proposition has a natural interpretation. 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. In a setting of complete interbank deposits (or, indeed, in any setting in which interbank deposits are increasing in the size of the negatively correlated regional liquidity shocks), the larger the anticipated regional liquidity shocks, the larger the dollar value of interbank deposits, and thus the larger the effect of contagion when it occurs. In another, somewhat looser, interpretation, note that the size of the regional liquidity shocks can be seen to be a measure of the level of intra-economy financial volatility. On this interpretation, the proposition above says that as financial volatility in the system increases, the damage caused by the failure of a financial institution also increases.

## 7 Discussion

We conclude with a few thoughts on the robustness of these results, and on potential extensions.

---

<sup>16</sup>Note that in a fully rational set-up, regional liquidity shocks are always anticipated ex ante. Since these shocks are negatively correlated, larger shocks imply larger interbank insurance levels, under the maintained assumption of complete interbank deposit insurance.

## 7.1 More regions? Aggregate liquidity shocks?

Our model extends naturally to more than two regions, and none of the results change in this extension. 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. The only significant change would be one of algebraic complexity in computing the level of interbank deposits, since imperfect negative correlation across regional liquidity shocks would lead to multi-party cross holdings of interbank deposits. The existence of interbank deposits, however complex, along with the seniority of institutional claims, shall create *ex interim* asymmetries amongst banks (some shall be debtors, others creditors) exactly as in the simpler two-region model discussed above. Thus, the contagious effect shall re-emerge in equilibrium exactly as before.

Adding aggregate liquidity shocks to our model creates a second source of bank failure in our model 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. While we do not deny that fully aggregate liquidity shocks may, indeed, emerge in an economy, we argue that the strategic nature of the game remains the same even in their presence. Therefore, we limit our attention to constant aggregate liquidity economies and show that contagion occurs *even in such economies*. Naturally, adding more sources for contagion will increase its occurrence.

## 7.2 Learning?

An important limitation of our model is that we abstract from social learning. 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. While this assumption seems fitting when discussing banks in New York and Oregon, for banks in neighboring regions, it seems less natural. Incorporating correlations amongst assets across the regions of our economy would introduce Bayes learning into our model and would complicate our arguments significantly. However, in recent work we have laid out techniques to analyze dynamic coordination games with social learning (Dasgupta 2001). Incorporating learning into a model similar to ours is a promising direction for future research. We conjecture that incorporating learning into our model would increase the occurrence of contagion, without modifying the main qualitative features outlined above.



## 8 Appendix

**Proof of Proposition 2:** 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  $\theta^*$ . 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  $\theta$ , with limits of integration given by functions of  $\theta^*$ , following the piecewise definition of  $\pi(\theta, n)$  above. Since the limits of integration are always linear with slope 1 in  $\theta^*$ , integrating over constant terms gives us final products that are independent of  $\theta^*$ . Thus, we can rewrite  $\Pi(\cdot)$  as

$$\int_{\theta^*+\epsilon(1-2r)}^{\theta^*+\epsilon} u\left[\frac{1-w-\frac{(1-(w+x))n_d+(x-D)}{r}-D}{(1-w-x)(1-n_d)}R(\theta_i)\right]d\theta - \int_{\theta^*-\epsilon}^{\theta^*+\epsilon(1-2n^d)} u\left[\frac{w+D+(1-w)r}{w+x+(1-w-x)n(\theta, \theta^*)}\right]d\theta + K \quad (13)$$

where  $K$  proxies for the terms that do not involve  $\theta^*$ . Taking the other parameters as given, we write:

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

and

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

and differentiate with respect to  $\theta^*$ :

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

Since the limits of integration, in each case are linear in  $\theta^*$ , 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 make the following set of observations:

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

Now, (1) and (3) imply that  $f(\theta, \theta^*)$  decreases in  $\theta^*$ . (1) and (4) imply that  $f(\theta, \theta^*)$  increases in  $\theta$ . (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  $\theta^*$ . (2) and (4) imply that  $g(\theta, \theta^*)$  increases in  $\theta$ . (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  $\theta^*$ . Thus, there is only one value of  $\theta^*$  that solves  $\Pi(\theta^*, \theta^*) = 0$ . ■

**Proof of Proposition 3** As always, we only present the proof for the static coordination game for the debtor bank's patient depositors. The proofs for all other static games are identical, with simpler payoffs. 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  $\theta$ . Then  $\frac{dn(\theta)}{d\theta} \in [-\frac{1}{2\epsilon}, \frac{1}{2\epsilon}]$*

**Proof:** At state  $\theta$ , 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  $\theta_i$ . Then, for this agent:

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

Differentiating relative to  $\theta$ , 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  $\theta \in [\theta_B, \theta_T]$   $\hat{\theta}(\theta) \leq \theta_B$  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

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

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  $\theta$  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  $\theta$  and  $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) < 0$  for some other  $\theta$  in  $[\theta_B, \theta_T]$ . Since  $\frac{\theta_T - \theta}{2\epsilon}$  is monotone in  $\theta$ , there is exactly one point, call it  $\theta_1$  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

$$\int_{\theta_B}^{\theta_1} \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) d\theta \geq \int_{\theta_B}^{\theta_2} \pi(\hat{\theta}(\theta), n(\theta)) d\theta \quad (14)$$

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  $\theta$ ,  $\pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) < 0$ . 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  $\theta_1$ ,

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

At  $\theta_2$ ,

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

Since  $\hat{\theta}(\theta_2) < \theta_1$ ,  $n(\theta_2) > \frac{\theta_T - \theta_1}{2\epsilon}$ . This completes the proof of the claim.

Now we shall use Claims (1) and (2) to demonstrate (14). Denote  $m(\theta) = \frac{\theta_T - \theta}{2\epsilon}$ . By a change of variables:

$$\begin{aligned} \int_{\theta_B}^{\theta_1} \pi(\theta, m(\theta)) d\theta &= \int_{m(\theta_1)}^{m(\theta_B)} \pi(\theta(m), m) \left| \frac{\partial \theta}{\partial m} \right| dm \\ \int_{\theta_B}^{\theta_2} \pi(\hat{\theta}(\theta), n(\theta)) d\theta &= \int_{\min[n(\theta): \theta \in [\theta_B, \theta_2]]}^{\max[n(\theta): \theta \in [\theta_B, \theta_2]]} \pi(\hat{\theta}(\theta(n)), n) \left| \frac{\partial \theta}{\partial n} \right| dn + \int_{\theta: \frac{\partial \theta}{\partial n} = 0} \pi(\hat{\theta}(\theta(n)), n) d\theta \end{aligned}$$

The second integral is smaller, because it is computed over a range that is smaller (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 (14).

Since  $m(\theta)$  declines faster than  $n(\theta)$ , and by Claim 2, we know that  $\theta_2 > \theta_1$ . Thus,

$$\int_{\theta_1}^{\theta_T} \pi(\theta, m(\theta)) d\theta \geq \int_{\theta_2}^{\theta_T} \pi(\theta, m(\theta)) d\theta > \int_{\theta_2}^{\theta_T} \pi(\hat{\theta}(\theta), n(\theta)) d\theta \quad (15)$$

We combine (14) and (15) 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  $\theta^*$  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  $\theta^*$ . Thus, if we could show that under any feasible beliefs  $n$  over the actions of other agents, there can be only one point  $\theta^*$  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.

Assume the contrary. Let  $\theta_H = \sup\{\theta_i : \Pi(\theta_i, n) = 0\}$ , and suppose there exist (possibly many) zeros of  $\Pi(\theta_i, n)$  that are strictly smaller than  $\theta_H$ . Let  $\theta_L = \sup\{\theta_i : \theta_i < \theta_H, \Pi(\theta_i, n) = 0\}$ . It is easy to see that  $\Pi(\theta_i, n) < 0$  for all  $\theta_i \in (\theta_L, \theta_H)$ .

Consider the case when  $\theta_H - \theta_L < 2\epsilon$ .<sup>17</sup> 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 simple change of variables  $\hat{\theta}(\theta) = \theta_L + \theta_H - \theta$ , we can re-write this as:

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

**Claim 3** *There are two parts:*

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

---

<sup>17</sup>The complementary case has an essentially identical, but simpler, proof, and is omitted.

**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  $\theta_i$  for which  $\Pi(\theta_i) < 0$  lie in  $(\theta_L, \theta_H)$ . 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  $\theta_L$  and  $\theta_H$ , 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,

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

where  $\theta_B = \theta_L + \epsilon$ , and  $\theta_T = \theta_H + \epsilon$ . It is easy to see that  $\int_{\theta_B}^{\theta_T} \pi(\theta, \frac{\theta_T - \theta}{2\epsilon}) d\theta \geq 0$ . When  $\hat{\theta} = \theta_B$ ,  $\theta = \theta_H - \epsilon$ . Thus, given the claim above, and given that  $\frac{\theta_T - \epsilon}{2\epsilon}$  decreases at the fastest feasible rate, we can say, for all  $\theta \in [\theta_B, \theta_T]$ ,  $n(\theta) \geq \frac{\theta_T - \epsilon}{2\epsilon}$ . We can now directly apply Lemma 2 to claim that  $\Pi(\theta_H, n) - \Pi(\theta_L, n) > 0$ , a contradiction. ■

**Proof of Proposition 5:** 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,

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

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

where  $K_1 = \int_{\theta^* + \epsilon l_1}^{\theta^* + \epsilon} u[1] d\theta$ . 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

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

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

where  $K_1$  is as before. Observe that since  $g > 0$ ,  $L_s(\theta^*) > L_f(\theta^*)$  for all  $\theta^*$ . 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  $\theta^*$ . 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 proposition 2 we know that  $L_s(\theta^*)$  is increasing in  $\theta^*$ , while  $R_s(\theta^*)$  is invariant with  $\theta^*$ . Thus, in order to make the indifference equations hold, we need to reduce  $\theta^*$  below  $\theta_{c,F}^*$ , and thus, we have just shown that  $\theta_{c,S}^* < \theta_{c,F}^*$ . ■

**Proof of Proposition 6:** We begin by proving two intermediate results:

**Lemma 3**  $r(1 - w) + (1 - r(1 - w) - w + x) \ln \left[ 1 - \frac{1 - w}{1 - w + x} r \right]$  is decreasing in  $x$  and positive over the permissible range of  $x$ .

**Proof:** Let

$$E(r, w, x) = r(1 - w) + (1 - r(1 - w) - w + x) \ln \left[ 1 - \frac{1 - w}{1 - w + x} r \right]$$

Write  $a = (1 - w)(1 - r)$  and  $b = (1 - w)$ . Observe that  $a = b(1 - r)$ . Now,

$$E(r, a, b, x) = (a + x) \ln \left[ 1 - \frac{br}{b + x} \right]$$

which simplifies to

$$E(a, b, x) = (a + x) \ln \left[ \frac{a + x}{b + x} \right]$$

We differentiate with respect to  $x$  to obtain

$$E'[a, b, x] = \frac{b - a}{b + x} + \ln \left[ \frac{a + x}{b + x} \right]$$

Is this expression always negative over the permissible range of  $x$ ? To investigate, we differentiate again with respect to  $x$ , and obtain

$$E''(a, b, x) = \frac{b - a}{b + x} \left[ \frac{1}{a + x} - \frac{1}{b + x} \right] > 0$$

Thus, if  $E' < 0$  for the maximal permissible  $x$ , we shall be done. But note that  $x < \min[w, (1 - w)(1 - r)] = \min[1 - b, a]$ . We check that for  $x = 1 - b$ ,  $E'$  is negative.

$$E'(a, b, 1 - b) = (b - a) + \ln(1 - (b - a))$$

Let  $y = b - a$ . This is then equivalent to

$$y > 1 - \exp(-y)$$

which is always true for  $y \geq 0$ . Thus, we have just shown that  $r(1-w) + (1-r(1-w) - w + x) \ln \left[ 1 - \frac{1-w}{1-w+x} r \right]$  is decreasing in  $x$ . To show that it is always positive, we simply need to check that it is positive at the highest global value of  $x$ , which is  $\frac{1}{2}$ . Note that  $x = \frac{1}{2}$  implies that  $w = \frac{1}{2}$  to keep everything well defined. Thus, our target expression reduces to

$$E(r, \frac{1}{2}, \frac{1}{2}) = \frac{r}{2} + (1 - \frac{r}{2}) \ln \left[ 1 - \frac{r}{2} \right]$$

which is clearly positive for all  $r \in (0, 1)$ . ■

**Lemma 4**  $\theta_d^*$  increases in  $x$ .

**Proof:** Observe that

$$\theta_d^* = \left[ \frac{r}{r + (1-r) \ln(1-r)} \right] \left[ \frac{(1-w) + (r(1-w) + w_x) \ln \left[ \frac{1}{r(1-w) + w + x} \right]}{(1-w-x)} \right]$$

Observe that the left term in the product is always positive and is not affected by  $x$ . Thus, to analyze the dependence of  $\theta_d^*$  on  $x$ , it is sufficient to examine the right term in the product. We differentiate this with respect to  $x$ , to obtain

$$\frac{[x - (1-r)(1-w)] - [1 + r(1-w)] \ln[r(1-w) + w + x]}{(1-w-x)^2}$$

Since the denominator is always positive, if we could show that the numerator is positive, then we shall be done. Observe that by differentiating the numerator with respect to  $x$ , we get

$$1 - \frac{1 + r(1-w)}{r(1-w) + w + x}$$

Since  $x < (1-r)(1-w)$ , the numerator is decreasing in  $x$ . Thus, if the numerator is positive for the maximal  $x$ , then it is positive for all  $x$ , and we are done. We check that the numerator is 0 for  $x = (1-r)(1-w)$ , and so we are done. ■

We are now ready to prove the main result. In the limit, the survivor of the debtor bank implies that  $n_d = 0$ , so that  $\hat{g}(\cdot) = x(\frac{\theta_d^* + U}{2})$ . Now, we can use the expressions for the equilibrium thresholds shown above and write:

$$ct(r, w, x) = \frac{[-x(\frac{\theta_d^* + U}{2})] r \ln \left[ 1 - \frac{1-w}{1-w+x} r \right]}{r(1-w) + (1-r(1-w) - w + x) \ln \left[ 1 - \frac{1-w}{1-w+x} r \right]}$$



The numerator of  $ct(x)$  can be rewritten as  $[(\frac{\theta_d^*+U}{2})]r(-x \ln [1 - \frac{1-w}{1-w+x}r])$ .  $\frac{\theta_d^*+U}{2}$  is increasing in  $\theta_d^*$ , which in turn is increasing in  $x$  by Lemma 4.  $-x \ln [1 - \frac{1-w}{1-w+x}r]$  is increasing in  $x$ . By Lemma 3 we know that the denominator decreases in  $x$ . Thus,  $ct(x)$  increases in  $x$ . ■

## References

- [1] Acharya, V. and T. Yorulmazer (2002), “Information Contagion and Inter-Bank Correlation in a Theory of Systemic Risk”, Mimeo, London Business School.
- [2] Allen, F. and D. Gale (2000), “Financial Contagion,” *Journal of Political Economy*, 108, 1-33.
- [3] Calomiris, C. and G. Gorton (1991), “The Origins of Banking Panics: Models, Facts, and Bank Regulation,” in G. Hubbard, ed. *Financial Markets and Financial Crises*, NBER Project Report, University of Chicago Press, Chicago and London, 109-73.
- [4] Calomiris, C. and C. Kahn (1991), “The Role of Demandable Debt in Structuring Optimal Banking Arrangements,” *American Economic Review*, 81, 497-513.
- [5] Chari, V. (1989), “Banking Without Deposit Insurance or Bank Panics: Lessons from a Model of the U.S. National Banking System,” *Federal Reserve Bank of Minneapolis Quarterly Review*, 13, 3-19.
- [6] Chen, Y. (1999), “Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities,” *Journal of Political Economy*, 107, 946-68.
- [7] Dasgupta, A. (2001), “Coordination, Learning, and Delay,” Mimeo, Yale University.
- [8] Diamond, D. and P. Dybvig (1983), “Bank Runs, Deposit Insurance, and Liquidity,” *Journal of Political Economy*, 91, 401-19.
- [9] Goldstein, I. and A. Pauzner (2000a), “Demand Deposit Contracts and the Probability of Bank Runs,” Mimeo, Tel Aviv University.
- [10] Goldstein, I. and A. Pauzner (2000b), “Contagion of Self-Fulfilling Financial Crises Due to Diversification of Investment Portfolios,” Mimeo, Tel Aviv University.
- [11] Kodres, L. and M. Pritsker (1998), “A Rational Expectations Model of Financial Contagion”, *Board of Governors of the Federal Reserve System: Finance and Economics Discussion Paper Series 98/48*, pages 57.

- [12] Kiyotaki, N. and J. Moore (2001), “Credit Chains,” Clarendon Lectures, Oxford, UK.
- [13] Kyle, A. and W. Xiong (2001), “Contagion as a Wealth Effect,” *Journal of Finance*, 56, 1401-1440.
- [14] Morris, S. and H. Shin (2000), “Global Games: Theory and Applications”, forthcoming in the *Volume of the Eighth World Congress of the Econometric Society*..
- [15] Rochet, J. and J. Tirole (1996) “Interbank Lending and Systemic Risk,” *Journal of Money, Credit, and Banking*, 28, 733-62.
- [16] Rochet, J. and X. Vives (2000), “Coordination Failure and the Lender of Last Resort,” Mimeo, Toulouse University.
- [17] Sprague, O. M. W. (1910), *History of Crises Under the National Banking System*, Washington, Government Printing Office.
- [18] Wicker, E. (2000), *Banking Panics of the Gilded Age*, Cambridge University Press, New York and Melbourne.