Financial Intermediation and Crises

Itay Goldstein

Wharton School, University of Pennsylvania

Risk Sharing and Bank Runs: Diamond and Dybvig (1983)

- Diamond and Dybvig provide a seminal model of financial intermediation and **bank runs**.
- Banks Create liquid claims on illiquid assets using **demand-deposit contracts**.
  - Enable investors with early liquidity needs to participate in long-term investments. Provide **risk sharing**.
  - Drawback: Contracts expose banks to panic-based bank runs.

Model (Extended)

- There are three periods (*0*, *1*, *2*), one good, and a continuum *[0,1]* of agents.
- Each agent is born at period *0* with an endowment of *1*.
- Consumption occurs only at periods *1* or *2*.
- Agents can be of two types:
    - Impatient (probability λ) – enjoys utility $u(c_1)$,
    - Patient (probability *1*-λ) – enjoys utility $u(c_1 + c_2)$.

- Types are i.i.d., privately revealed to agents at the beginning of period *1*.

- Agents are highly risk averse. Their relative risk aversion coefficient:

$$-\frac{cu''(c)}{u'(c)} > 1 \text{ for any } c \geq 1.$$

  - This implies that $cu'(c)$ is decreasing in c for $c \geq 1$, and hence $cu'(c) < u'(1)$ for $c > 1$.

  - Assume $u(0) = 0$.

- Agents have access to the following technology:
  - *1* unit of input at period *0* generates *1* unit of output at period *1* or *R* units at period *2* with probability $p(\theta)$.
  - θ is distributed uniformly over *[0,1]*. It is revealed at period *2*.
  - $p(\theta)$ is increasing in θ.
  - The technology yields (on average) higher returns in the long run:

$$E_\theta[p(\theta)]u(R) > u(1).$$

Autarky

- In autarky, impatient agents consume in period 1, while patient agents wait till period 2. The expected utility is then:

$$\lambda u(1) + (1 - \lambda) u(R) E_\theta[p(\theta)]$$

- Because agents are risk averse, there is a potential gain from transferring consumption from impatient agents to patient agents, and letting impatient agents benefit from the fruits of the long-term technology.

- We now derive the first-best and see how it can be implemented.

First-Best Allocation (if types were verifiable)

- A social planner verifies types and allocates consumptions.
- Period-1 consumption of impatient agents: c_1.
- Period-2 consumption of patient agents is the remaining resources: $c_2 = \frac{(1-\lambda c_1)}{1-\lambda} R$ (with probability $p(\theta)$).
- Planner sets c_1 to maximize expected utility:

$$\lambda u(c_1) + (1-\lambda) u\left(\frac{(1-\lambda c_1)}{1-\lambda} R\right) E_\theta[p(\theta)]$$

- First order condition:

$$u'(c_1^{FB}) = Ru'\left(\frac{(1-\lambda c_1^{FB})}{1-\lambda}R\right)E_\theta[p(\theta)]$$

- Suppose that $\mathrm{c}_1^{\mathrm{FB}} = 1$: $\mathrm{u}'(1) > Ru'(\mathrm{R})\mathrm{E}_\theta[\mathrm{p}(\theta)]$.
- Since the LHS is decreasing and the RHS is increasing in c_1^{FB}, we get that: $c_1^{FB} > 1$.
- The social planner achieves risk sharing by liquidating a larger portion of the long-term technology and giving it to impatient agents. The benefit of risk sharing outweighs the cost of lost output.

The Role of Banks

- The main insight of Diamond and Dybvig is that banks can replicate the first-best allocation with demand-deposit contracts.
  - Hence, they overcome the fact that types are not verifiable.
- Banks offer a short-term payment r_1 to every agent who claims to be impatient.
- By setting $r_1 = c_1^{FB}$, they can achieve the first-best allocation, as long as the incentive compatibility constraint holds:

$$u(c_1^{FB}) \leq u\left(\frac{(1-\lambda c_1^{FB})}{1-\lambda}R\right)E_\theta[p(\theta)]$$

- Yet, things are not so simple, as one has to think carefully about the mechanic details of how banks serve agents and the resulting equilibria.
- Suppose that banks follow a sequential service constraint:
    - They pay r_1 to agents who demand early withdrawal as long as they have resources.
    - If too many agents come and they run out of resources, they go bankrupt, and remaining agents get no payment.
- Impatient agents demand early withdrawal since they have no choice. Patient agent have to consider the following payoff matrix:

Period	$n < 1/r_1$	$n \geq 1/r_1$
1	r_1	$\begin{cases} r_1 & prob \quad \dfrac{1}{nr_1} \\ 0 & prob\ 1 - \dfrac{1}{nr_1} \end{cases}$
2	$\begin{cases} \dfrac{(1 - nr_1)}{1 - n} R & prob \quad p(\theta) \\ 0 & prob\ 1 - p(\theta) \end{cases}$	0

Here, n is the proportion of agents (patient and impatient who demand early withdrawal.

Multiple Equilibria

- Assuming that the incentive compatibility condition holds, there are at least two Nash equilibria here:
  - **Good equilibrium:** only impatient agents demand early withdrawal.
    - Clear improvement over autarky. First-best is achieved.
  - **Bad equilibrium:** all agents demand early withdrawal. **Bank Run** occurs.
    - Inferior outcome to autarky. No one gets access to long-term technology and resources are allocated unequally.

Source and Nature of Bank Runs

- Bank runs occur because of **strategic complementarities** among agents. They want to do what other agents do.
  - When everyone runs on the bank, this depletes the bank's resources, and makes running the optimal choice.

- As a result, runs are **panic-based**: They occur as a result of the **self-fulfilling beliefs** that other depositors are going to run.

- Moreover, here, they are **unrelated to fundamentals**.
  - Some tend to attribute them to sunspots.

Problem in the Model

- The model provides no tools to determine which equilibrium is more likely to occur.

- If the probability of runs is non-negligible:
  - ‘Optimal’ contract may be not optimal. Demand deposit contract may be not even desirable.

- If the probability depends on contract:
  - It would affect the optimal contract. Demand deposit contract may be even *more* undesirable.

Solutions to Fragility – Suspension of Convertibility:

- Suppose that the bank announces that after λ depositors withdraw in period 1, no one else gets money in this period.
- The good equilibrium becomes the unique equilibrium.
- Patient agents know that no matter what others do, they are guaranteed to get $u\left(\frac{\left(1-\lambda c_1^{FB}\right)}{1-\lambda}R\right)E_\theta[p(\theta)] > u(c_1^{FB})$.
- Hence, the run is prevented without even triggering suspension.
- **Problem:** What if the number of impatient agents is not known? Suspension of convertibility may severely hurt impatient agents.

Solutions to Fragility – Deposit Insurance:

- Suppose that the government provides insurance to the bank in case of excess withdrawals.
    - To maintain the assumption of 'closed' economy, suppose that the government obtains this amount by taxing depositors.
- Again, the good equilibrium becomes the unique equilibrium.
    - Patient agents know that the withdrawal by others is not going to harm their long-term return.
- **Problem:** Deposit insurance might generate moral hazard: Banks make too risky investments or set deposit rate too high.

The Global-Games Approach: Morris and Shin (1998)

- The global-games approach – based on Carlsson and van Damme (1993) – enables us to derive a unique equilibrium in a model with strategic complementarities and thus overcome the problems associated with multiplicity of equilibria (discussed above).
- The approach assumes **lack of common knowledge** obtained by assuming that agents observe slightly noisy signals of the fundamentals of the economy.
- The classic illustration is by Morris and Shin (1998).

A Model of Currency Attacks

- There is a continuum of speculators *[0,1]* and a government.
- The exchange rate without intervention is $f(\theta)$, where $f'(\theta) > 0$, and θ, the fundamental of the economy, is uniformly distributed between *0* and *1*.
- The government maintains the exchange rate at an over-appreciated level (due to reasons outside the model): $e^* > f(\theta), \forall\theta$.
- Speculators may choose to attack the currency.

- The cost of attack is t (transaction cost).
- The benefit in case the government abandons is $e^* - f(\theta)$.
  - In this case, speculators make a speculative gain.

- The government's payoff from maintaining is: $v - c(\alpha, \theta)$.
  - v can be thought of as reputation gain.
    - $c(\alpha, \theta)$ is increasing in α (proportion of attackers) and decreasing in θ: Cost increases in loss of reserves and decreases in fundamentals.

Equilibria under Perfect Information

- Suppose that all speculators (and the government) have perfect information about the fundamental θ.

- Define extreme values of θ, $\underline{\theta}$ and $\overline{\theta}$: $1 > \overline{\theta} > \underline{\theta} > 0$, such that:

  - $c\left(0, \underline{\theta}\right) = v$.

  - $e^* - f\left(\overline{\theta}\right) = t$.

  - Below $\underline{\theta}$, the government always abandons. Above $\overline{\theta}$, attack never pays off.

- Three ranges of the fundamentals:
    - When $\theta < \underline{\theta}$, unique equilibrium: all speculators attack.
    - When $\theta > \overline{\theta}$, unique equilibrium: no speculator attacks.
    - When $\overline{\theta} > \theta > \underline{\theta}$, multiple equilibria: Either all speculators attack or no speculator attacks (for this, assume $c(1,1) > v$).
- As in Diamond and Dybvig, the problem of multiplicity comes from strategic complementarities: when others attack, the government is more likely to abandon, increasing the incentive to attack.

Introducing Imperfect Information

- Suppose that speculator i observes $\theta_i = \theta + \varepsilon_i$, where ε_i is uniformly distributed between $-\varepsilon$ and ε. (Government has perfect information.)

- Speculators choose whether to attack or not based on their signals.

- The key aspect is that because they only observe imperfect signals, they must take into account what others will do at other signals.

- This will 'connect' the different fundamentals and determine optimal action at each.

Definitions

- Payoff from attack as function of fundamental and aggregate attack:

$$h\big(\theta,\alpha(\theta)\big) = \begin{cases} e^* - f(\theta) - t & if \quad \alpha(\theta) > a(\theta) \\ -t & if \quad \alpha(\theta) \le a(\theta) \end{cases},$$

  where $c(a(\theta),\theta) = v$.

- Payoff as a function of the signal and aggregate attack:

$$V\big(\theta_i,\alpha(\theta)\big) = \frac{1}{2\varepsilon}\int_{\theta_i-\varepsilon}^{\theta_i+\varepsilon} h\big(\theta,\alpha(\theta)\big)d\theta$$

- Threshold strategy characterized by θ' is a strategy where the speculator attacks at all signals below θ' and does not attack at all signals above θ'.
  - Aggregate attack when speculators follow threshold θ':

$$\alpha(\theta, \theta') = \begin{cases} 0 & if \quad \theta > \theta' + \varepsilon \\ \dfrac{\theta' + \varepsilon - \theta}{2\varepsilon} & if \quad \theta' - \varepsilon \leq \theta \leq \theta' + \varepsilon \\ 0 & if \quad \theta < \theta' - \varepsilon \end{cases}$$

- We will show that there is a unique threshold equilibrium and no non-threshold equilibria that satisfy the Bayesian-Nash definition.

Existence and Uniqueness of Threshold Equilibrium

- Let us focus on the incentive to attack at the threshold:
    - Function $V\big(\theta', \alpha(\theta, \theta')\big)$ is monotonically decreasing in θ'; positive for low θ' and negative for high θ'.
    - Hence, there is a unique θ^* that satisfies $V\big(\theta^*, \alpha(\theta, \theta^*)\big) = 0$.
    - This is the only candidate for a threshold equilibrium, as in such an equilibrium, at the threshold, speculators ought to be indifferent between attacking and not attacking.

- To show that acting according to threshold θ^* is indeed an equilibrium, we need to show that speculators with lower signals wish to attack and those with higher signals do not wish to attack.
    - This holds because: $V\left(\theta_i, \alpha(\theta, \theta^*)\right) > V\left(\theta^*, \alpha(\theta, \theta^*)\right) = 0$, $\forall \theta_i < \theta^*$, due to the direct effect of fundamentals (lower fundamental, higher profit and higher probability of abandoning) and that of the attack of others (lower fundamental, more people attack and higher probability of abandoning).
    - Similarly, $V\left(\theta_i, \alpha(\theta, \theta^*)\right) < V\left(\theta^*, \alpha(\theta, \theta^*)\right) = 0$, $\forall \theta_i > \theta^*$,

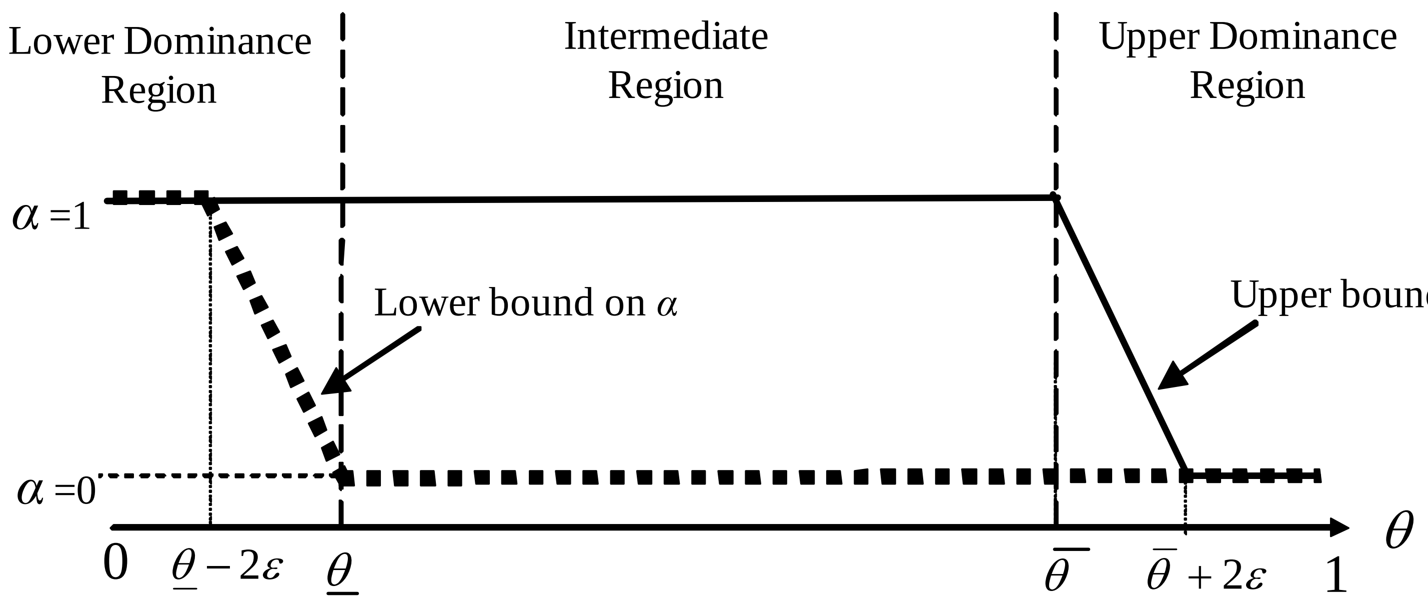
Ruling out Non-Threshold Equilibria

- These are equilibria where agents do not act according to a threshold strategy.
- By contradiction, assume such an equilibrium and suppose that speculators attack at signals above θ^*; denote the highest such signal as θ'^* (we know it is below 1 because of upper dominance region).
- Denote the equilibrium attack as $\alpha'(\theta)$, then due to indifference at a switching point: $V\left(\theta'^*, \alpha'(\theta)\right) = 0$.
- We know that $\alpha'(\theta) \leq \alpha(\theta, \theta'^*)$.

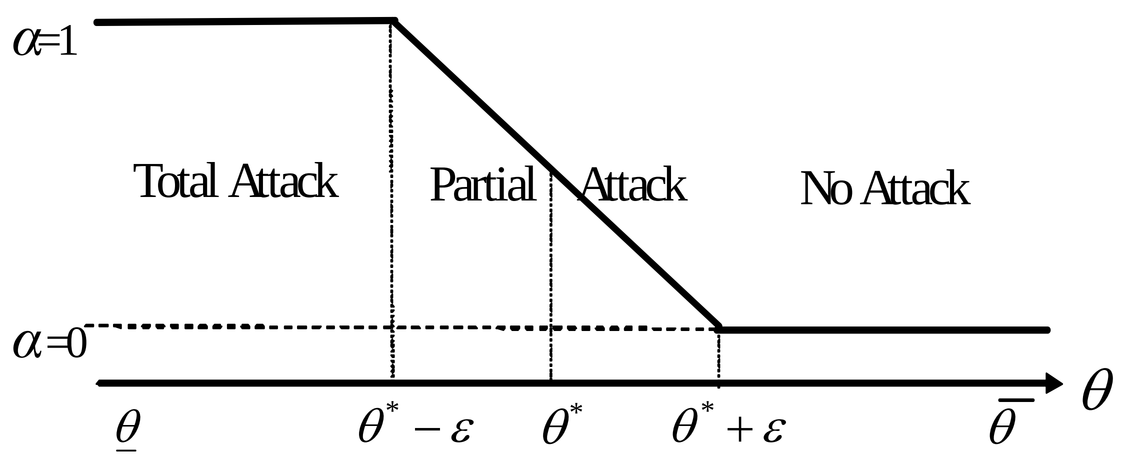
- Then, due to strategic complementarities: $V\left(\theta'^* , \alpha(\theta, \theta'^*)\right) \geq 0$.
- But, this is in contradiction with $V\left(\theta^*, \alpha(\theta, \theta^*)\right) = 0$, since θ'^* is above θ^* and function $V\left(\theta', \alpha(\theta, \theta')\right)$ is monotonically decreasing in θ'.
- Hence, speculators do not attack at signals above θ^*.
- Similarly, one can show that they always attack at signals below θ^*.
- This rules out equilibria that are different than a threshold equilibrium, and establishes the threshold equilibrium based on θ^* as the unique equilibrium of the game.

Some Intuition

- These are the bounds on the proportion of attack imposed by the dominance regions:



- These bounds can be shifted closer together by iterative elimination of dominated strategies.
- The result is the equilibrium that we found:



Important:

- Although θ uniquely determines α, attacks are still driven by bad expectations, i.e., still panic-based:
  - In the intermediate region speculators attack because they believe others do so.
  - θ acts like a coordination device for agents' beliefs.
- A crucial point: θ is not just a sunspot, but rather a payoff-relevant variable.
  - Agents are obliged to act according to θ.

Why Is This Equilibrium Interesting?

- **First**, reconciles panic-based approach with empirical evidence that fundamentals are linked to crises.
- **Second**, panic-based approach generates empirical implications.
  - Here, the probability of a crisis is pinned down by the value of θ^*, which depends on variables like t, v, etc.
- **Third**, once the probability of crises is known, one can use the model for policy implications.
- **Fourth**, captures the notion of strategic risk, which is missing from the perfect-information version.

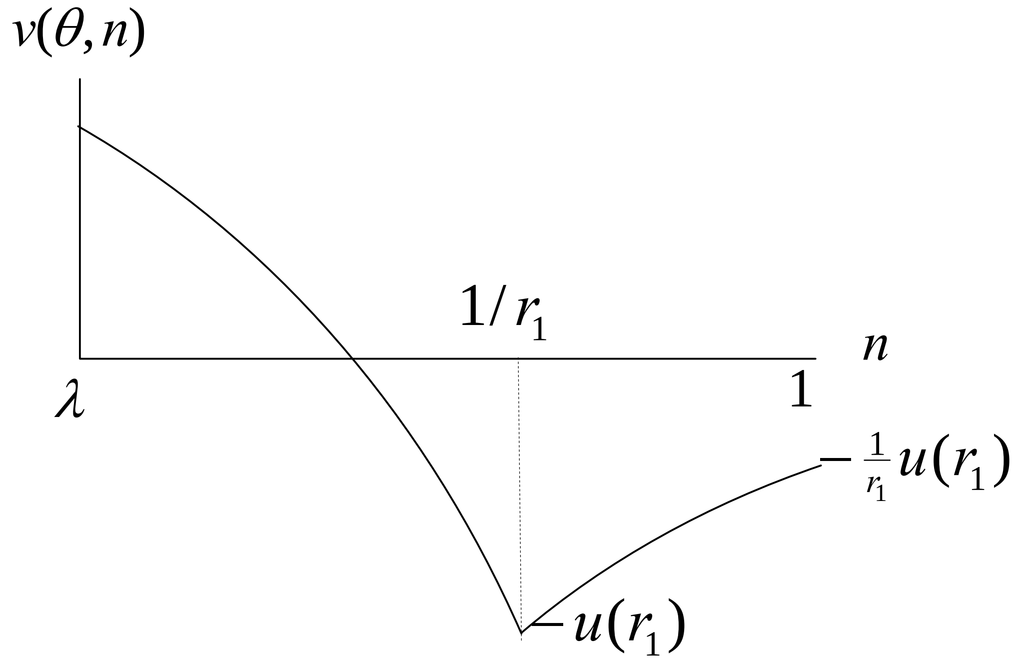
Back to Bank Runs: Goldstein and Pauzner (2005)

- Use global-games approach to address the fundamental issues in the Diamond-Dybvig model.
- But, the Diamond-Dybvig model violates the basic assumptions in the global-games approach. It does not satisfy **global strategic complementarities**.
    - Derive new proof technique that overcomes this problem.
- Once a unique equilibrium is obtained, study how the probability of a bank run is affected by the banking contract, and what is the optimal demand-deposit contract once this is taken into account.

Reminder, Payoff Structure

Period	$n < 1/r_1$	$n \geq 1/r_1$
1	r_1	$\begin{cases} r_1 & prob \quad \dfrac{1}{nr_1} \\ 0 & prob\ 1 - \dfrac{1}{nr_1} \end{cases}$
2	$\begin{cases} \dfrac{(1-nr_1)}{1-n}R & prob \quad p(\theta) \\ 0 & prob\ 1-p(\theta) \end{cases}$	0

- Global strategic complementarities do not hold:
    - An agent's incentive to run is highest when n=$1/r_1$ rather than when $n = 1$.
- Graphically:



- The proof of uniqueness builds on **one-sided strategic complementarities**:
    - v is monotonically decreasing whenever it is positive
- which implies **single crossing**:
    - v crosses zero only once.
- Show uniqueness by:
    - Showing that there exists a unique threshold equilibrium.
    - Showing that every equilibrium must be a threshold equilibrium.

The Demand-Deposit Contract and the Viability of Banks

- We can now characterize the threshold as a function of the rate offered by banks for early withdrawals. At the limit, as ε approaches zero, $\theta^*(r_1)$ is defined by:

$$\int_{n=\lambda}^{1/r_1} u(r_1) + \int_{n=1/r_1}^{1} \frac{1}{nr_1} u(r_1) = \int_{n=\lambda}^{1/r_1} p(\theta^*) u\left(\frac{(1-nr_1)}{1-n} R\right)$$

  - At the threshold, a patient agent is indifferent.
  - His belief at this point is that the proportion of other patient agents who run is uniformly distributed. Effectively, there is no fundamental uncertainty (only strategic uncertainty).

- Analyzing the threshold $\theta^*(r_1)$ with the implicit function theorem, we can see that it is increasing in r_1.
  - The bank becomes more vulnerable to bank runs when it offers more risk sharing.
- Intuition:
  - With a higher r_1 the incentive of agents to withdraw early is higher.
  - Moreover, other agents are more likely to withdraw at period 1, so the agent assesses a higher probability for a bank run.

Finding the optimal r_1

- The bank chooses r_1 to maximize the expected utility of agents:

$$\lim_{\varepsilon \to 0} EU(r_1) = \int_0^{\theta^*(r_1)} \frac{1}{r_1} u(r_1) d\theta$$

$$+ \int_{\theta^*(r_1)}^{1} \lambda u(r_1) + (1-\lambda) p(\theta) u\left(\frac{(1-\lambda r_1)}{1-\lambda} R\right) d\theta$$

- Now, the bank has to consider the effect that an increase in r_1 has on risk sharing and on the expected costs of bank runs.

- Main question: Are demand deposit contracts still desirable?

- Result: If $\underline{\theta}(1)$ is not too large, the optimal r_1 must be larger than 1.
- Increasing r_1 slightly above 1 generates one benefit and two costs:
  - **Benefit:** Risk sharing among agents.
    - Benefit is of first-order significance: Gains from risk sharing are maximal at r_1=1.
  - **Cost I:** Increase in the probability of bank runs beyond $\underline{\theta}(1)$.
    - Cost is of second order: Liquidation at $\underline{\theta}(1)$ is almost harmless.

- **Cost II:** Increase in the welfare loss resulting from bank runs below $\underline{\theta}(1)$.
  - Cost is small when $\underline{\theta}(1)$ is not too large.

- Hence, the optimal r_1 generates panic-based bank runs.

- But, the optimal r_1 is lower than c_1^{FB}.
  - Hence, the demand-deposit contract leaves some unexploited benefits of risk sharing in order to reduce fragility.
  - To see this, let us inspect the first order condition for r_1:

$$\lambda \int_{\theta^*(r_1)}^{1} u'(r_1) - p(\theta) R u'\left(\frac{(1-\lambda r_1)}{1-\lambda} R\right) d\theta =$$

$$\frac{\partial \theta^*(r_1)}{\partial r_1}\left(\lambda u(r_1) + (1-\lambda) p\big(\theta^*(r_1)\big) u\left(\frac{(1-\lambda r_1)}{1-\lambda} R\right) - \frac{1}{r_1} u(r_1)\right)$$

$$+ \int_{0}^{\theta^*(r_1)} \frac{u(r_1) - r_1 u'(r_1)}{r_1^2} d\theta$$

- LHS: marginal benefit from risk sharing. RHS: marginal cost of bank runs.
- Since marginal cost of bank runs is positive, and since marginal benefit is decreasing in r_1: The optimal r_1 is lower than c_1^{FB}.

Contagion

- One of the most striking features of financial Crises is that they spread across countries/institutions.
- Several leading explanations have been offered:
  - Information.
  - Interbank Connections.
  - Investors’ portfolios readjustments.
  - Behavioral explanations.

Interbank Connections: Allen and Gale (2000)

- There are three dates: 0, 1, and 2; one good.
- Investment technology:
    - **Short term**: One unit invested in $t=0$ yields one unit in $t=1$.
    - **Long term**: One unit invested in $t=0$ yields R in $t=2$, or r in $t=1$; $0<r<1<R$.
- There are four different regions: A, B, C, and D.
    - Each region has a continuum [0,1] of agents, who might face liquidity shocks, as in Diamond and Dybvig.

- Utility is given by:

$$U(c_1, c_2) = \begin{cases} u(c_1)\ prob & \omega \\ u(c_2)\ prob\ 1-\omega \end{cases}$$

- The probability of a liquidity shock varies from region to region; there are two equally likely states:

TABLE 1

REGIONAL LIQUIDITY SHOCKS

	A	B	C	D
S_1	ω_H	ω_L	ω_H	ω_L
S_2	ω_L	ω_H	ω_L	ω_H

Optimal Risk Sharing

- Denote $\gamma = (\omega_H + \omega_L)/2$.

- Planner maximizes:

$$\gamma u(c_1) + (1-\gamma)u\left(\frac{1-\gamma c_1}{1-\gamma}R\right)$$

- Hence,

$$u'(c_1) = u'\left(\frac{1-\gamma c_1}{1-\gamma}R\right)R$$

- Achieved by investing c_1 in short asset and $\frac{1-\gamma c_1}{1-\gamma}$ in long asset.
- First-best allocation satisfies incentive-compatibility constraint
    - Thus, first-best can be achieved even if types are not observable.
- The allocation ignores division to regions, and resources move across them to absorb liquidity needs.
- In particular, the planner will shift resources across regions.
    - In state 1, $(\omega_H - \gamma)c_1$ moves from B and D to A and C in *t=1*, and $(\omega_H - \gamma)c_2$ moves from A and C to B and D in *t=2*.

Decentralization

- In each region, consumers deposit their endowments in banks, who offer demand deposit contracts.
- Banks hold deposits in banks of other regions. Suppose the market is incomplete:

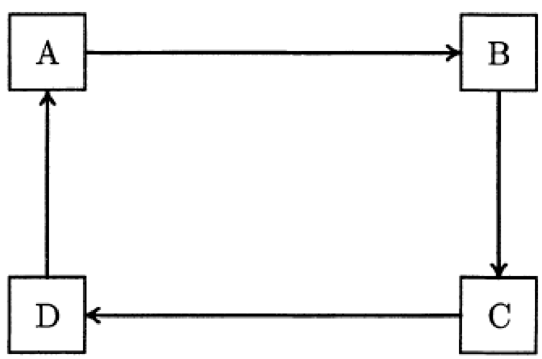


FIG. 2.—Incomplete market structure

- How can banks achieve the first best?
  - They make investments and promise returns as the planner.
  - They hold deposits of $\omega_H - \gamma$ at banks at the adjacent region.
  - In *t=1* banks in regions with high liquidity needs liquidate the deposits at banks in regions with low liquidity needs.
  - In *t=2* banks in regions with low liquidity needs liquidate the deposits at banks in regions with high liquidity needs.
- The fact that banks with low liquidity needs hold deposits in banks with high liquidity needs and vice versa guarantees efficient allocation.

Fragility

- Assume the same allocation as before, but a new state is possible:

TABLE 2

REGIONAL LIQUIDITY SHOCKS WITH PERTURBATION

	A	B	C	D
S_1	ω_H	ω_L	ω_H	ω_L
S_2	ω_L	ω_H	ω_L	ω_H
$\bar{S}$	$\gamma + \epsilon$	γ	γ	γ

- The new state is assigned probability zero; in it, aggregate demand for liquidity requires liquidation of some long-term assets.

- Assume that deposits are liquidated before long-term assets:

$$1 < \frac{c_2}{c_1} < \frac{R}{r}$$

- Banks start liquidating deposits in each other, and banks in region A liquidate some long-term assets.
- If aggregate liquidity shock is large enough, banks in region A must go bankrupt:
  - They liquidate long-term assets to pay early withdrawals, and cannot pay enough to patient investors, who then decide to run.

- If liquidation value is sufficiently low, banks in region D will also go bankrupt.
    - The value of their deposits in region A is low, so they liquidate long-term assets and trigger a run.
- By induction, banks in regions B and C will also go bankrupt.
- Overall, the failure of banks in region A, triggers a failure of region D, which triggers a failure of Region C, which triggers a failure of region B.

Interbank Structures that Reduce Fragility

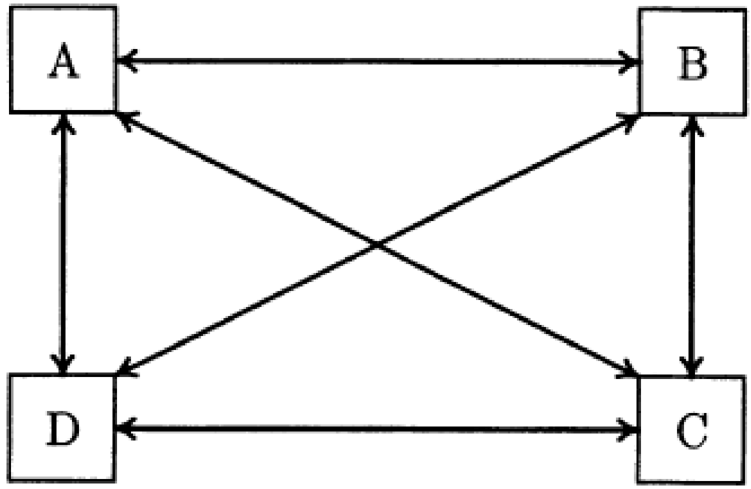


Fig. 1.—Complete market structure

- No bank depends strongly on banks in region A. The damage is spread out evenly, and not big enough to fail other regions.

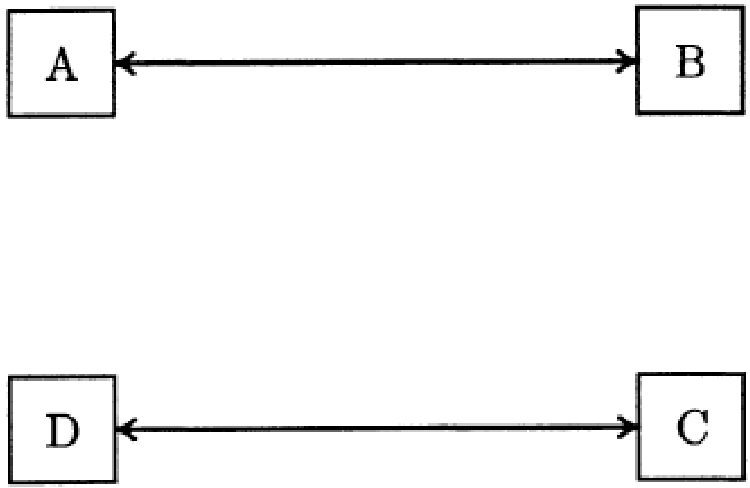


FIG. 3.—Disconnected incomplete market structure

- Failures are limited to regions A and B.
- Overall, the link between market completeness and fragility of the system is non-monotone.

The Monitoring Role of Financial Intermediaries: Diamond (1984)

- Under asymmetric information (borrower knows the return but lender doesn't), financing contracts require a deadweight loss:
  - The borrower should be punished when he doesn't pay a certain amount, for him to have an incentive to pay back.
  - This is costly when the borrower really cannot pay.
- Monitoring by the lender to verify returns may improve efficiency.
  - As long as monitoring costs are not too high.

- When there are many lenders, there is costly replication of monitoring efforts. Efficiency can be improved with **delegated monitoring**:
  - One financial intermediary monitors on behalf of all lenders.
- But, this solution is too naïve. The financial intermediary must be provided with adequate incentives as well.
  - Deadweight loss is required to incentivize him to pay.
- So when is financial intermediation a viable option?
  - **Answer:** when the financial intermediary holds a diversified portfolio. Then, the probability of not being able to pay is low.

The Model

- There are N entrepreneurs. Each one has access to investment technology that requires investment of 1 today, but has no wealth.
- The competitive interest rate in the economy is R, and the expected return on the investment technology $E(\tilde{y}) > R + K$. (K will be specified later.)
- The return $\tilde{y}$ is distributed between 0 and $\overline{y}$ and observed only by the entrepreneur.
- There are many lenders; each one has $1/m$ in wealth, and so each entrepreneur needs to get financing from $m>1$ lenders.

- The aggregate payment paid by the entrepreneur to the lenders is z.
- The entrepreneur can always claim a low y (0), and pay a low z (0), keeping y-z (y) to himself.
- In order to make the entrepreneur pay and enable financing to take place, we impose a non-pecuniary penalty (not enjoyed by the lender) on the entrepreneur when paying less than a certain amount.
    - This can be interpreted as cost of spending time in bankruptcy proceedings, cost of searching for a new job, etc.
    - This has similarities with deadweight cost in Townsend (1979), which was the cost of lender inspecting the return.

Optimal Contract

- The non-pecuniary penalty on the entrepreneur is $\phi(z)$.
- The optimal contract with penalties $\phi^*(z)$ solves:

$$\max_{\phi(z)} E_{\tilde{y}}[\tilde{y} - z - \phi(z)]$$

$$\text{Subject to } z \in \text{argmax}_{z\in[0,y]}\, y - z - \phi(z),$$

$$E_{\tilde{y}}\left[\text{argmax}_{z\in[0,y]}\, \tilde{y} - z - \phi(z)\right] \geq R.$$

- The assumption is that if the entrepreneur is indifferent between different z's, he chooses the one preferred by the lender.

- Under these assumptions, it can be shown that the optimal contract is a debt contract, given by:

$$\phi^*(z) = max(h - z, 0)$$

where h is the solution to:

$$\Big(P(\tilde{y} < h) \cdot E_{\tilde{y}}(\tilde{y}|y < h)\Big) + (P(\tilde{y} \geq h) \cdot h) = R.$$

- Here, h is the face value of the debt. The entrepreneur pays z=h as long as $y \geq h$.
  - He incurs a total loss $(z + \phi(z))$ of h.

- He has no reason to deviate to a lower payment, because the total loss will remain h.

- The entrepreneur pays z=y when $y < h$.
  - He cannot pay more.
  - He incurs a total loss of h, and has no incentive to deviate to a lower payment because the total loss will remain the same.
- The level of h is set such that the lenders get their required return.
- There is a deadweight loss of:

$$P(\tilde{y} < h) \cdot E_{\tilde{y}}(h - \tilde{y}|y < h)$$

Monitoring

- The deadweight loss involved in providing incentives to the entrepreneur to pay back can be avoided if lenders monitored the entrepreneur and verified his returns.
- Suppose that each lender can pay a fixed cost K and then know the return of the project.
  - Unlike in Townsend (1979), here the monitoring cost is incurred before the return on the project is realized.
- With m lenders, each one has to incur the monitoring cost, so this is worthwhile only as long as the deadweight loss is above mK.

Delegated Monitoring

- Another possibility is to let one financial intermediary monitor on behalf of all m lenders, reducing the monitoring cost to K.
- However, the problem is that the financial intermediary, just like the entrepreneur, must have proper incentives in place to find it worthwhile to pay back to the lenders.
- This leads to a deadweight loss D, and implies that delegated monitoring is viable only when:

$$K + D \leq min\left[P(\tilde{y} < h) \cdot E_{\tilde{y}}(h - \tilde{y}|y < h), (m \cdot K)\right]$$

Delegated Monitoring with One Entrepreneur

- In the case where the financial intermediary provides capital to one entrepreneur, financial intermediation is not viable.
- Suppose that the financial intermediary monitors, and obtains a payment $g(y) \leq y$ from the entrepreneur.
- Providing an incentive to the financial intermediary to pay to the lenders requires a debt contract with face value h_1 generating a deadweight loss of:

$$D = P(\tilde{g} < h_1) \cdot E_{\tilde{g}}(h_1 - \tilde{g} | g < h_1),$$

where h_1 is set to provide the lenders adequate return:

$$\left(P(\tilde{g} < h_1) \cdot E_{\tilde{g}}(\tilde{g}|g < h_1)\right) + (P(\tilde{g} \geq h_1) \cdot h_1) = R.$$

- Because $g(y) \leq y$, we get that $h_1 \geq h$, and thus the deadweight loss involved in providing incentives to the financial intermediary is greater than that needed to provide incentives to the entrepreneur.
- Adding this to the fact that the financial intermediary needs to pay a monitoring cost, it is clear that financial intermediation is not viable:

$$K + D \geq K + P(\tilde{y} < h) \cdot E_{\tilde{y}}(h - \tilde{y}|y < h)$$

$$> P(\tilde{y} < h) \cdot E_{\tilde{y}}(h - \tilde{y}|y < h).$$

Delegated Monitoring with Multiple Entrepreneurs

- The way to make financial intermediation work is to have the financial intermediary invest in multiple projects.
- Suppose he raises capital from mN lenders and uses it to finance N entrepreneurs with identically independently distributed returns.
- The advantage is coming from the fact that the intermediary's portfolio is now diversified, making it less likely that his return will be very low, and thus reducing the deadweight cost associated with incentive provision.

- To illustrate this, let us consider again the delegation cost with one entrepreneur:

$$D_1 = P(\widetilde{g_1} < h_1) \cdot E_{\widetilde{g_1}}(h_1 - \widetilde{g_1} | g_1 < h_1)$$

- Now, suppose that the intermediary works with two independent identical entrepreneurs, and that the face value of the debt doubles, then we get that the cost of delegation per entrepreneur is:

$$D_2 = \frac{1}{2} P(\widetilde{g_1} + \widetilde{g_2} < 2h_1)$$

$$\cdot E_{\widetilde{g_1}} E_{\widetilde{g_2}}(2h_1 - \widetilde{g_1} - \widetilde{g_2} | g_1 + g_2 < 2h_1)$$

- Using statistical rules, we can see that $D_2 < D_1$.
  - This is the usual effect of diversification.
- The result strengthens by the fact that, with two entrepreneurs, the face value of the debt will be even lower than $2h_1$ (because less is needed to provide adequate return to the lenders).
- The paper shows that with a very large number of independent projects, the delegation cost approaches zero.
  - This is a direct consequence of the law of large numbers.
- The results of viable intermediation extend also to the case with imperfectly correlated returns across the entrepreneurs.

Credit Frictions: Holmstrom and Tirole (1997)

- There is a continuum of firms with access to the same investment technology and different amounts of capital A.
- The distribution of assets across firms is described by the cumulative distribution function $G(A)$.
- The investment required is I, so a firm needs to raise I-A in external resources. The return is either 0 or R, and the probability depends on the type of project that the firm chooses.
- The firm may choose a lower type to enjoy private benefits.

Project	Good	Bad (low private benefit)	Bad (high private benefit)
Private benefit	0	b	B
Probability of success	p_H	p_L	p_L

- The rate of return demanded by investors is denoted as γ, which can either be fixed or coming from a supply function $S(\gamma)$.
- The assumption is that only the good project is viable:

$$p_H R - \gamma I > 0 > p_L R - \gamma I + B.$$

- The incentive of the firm to choose the good project will depend on how much “skin in the game” it has.
- Hence, it would be easier to finance firms with large assets A, since they are more likely to internalize the monetary benefit and choose the good project.

Financial Intermediaries

- In addition to investors who demand a rate of return γ, there are financial intermediaries, who can monitor the firm.
- Monitoring is assumed to prevent the firm from taking a B project, hence reducing the opportunity cost of the firm from B to b.
- Monitoring yields a private cost of c to the financial intermediary.
- Intermediary capital K_m will be important to provide incentives to the intermediary to monitor the firm (the Diamond solution of diversification is not considered here).

Direct Finance

- Consider a contract where the firm invests A, the investor invests I-A, no one gets anything if the project fails, and in case of success the firm gets R_f and the investor gets R_u:

$$R_f + R_u = R$$

- A necessary condition is that the firm has an incentive to choose the good project:

$$p_H R_f \geq p_L R_f + B.$$

- Denoting $\Delta p = p_H - p_L$, we get the incentive compatibility constraint:

$$R_f \geq B/\Delta p$$

- This implies that the maximum amount that can be promised to the investors (the **pledgeable expected income**) is:

$$p_H(R - B/\Delta p)$$

- Due to the participation constraint:

$$\gamma(I - A) \leq p_H(R - B/\Delta p)$$

- This puts a financing constraint on the firm that depends on how much internal capital it has.
- Defining

$$\overline{A}(\gamma) = I - p_H/\gamma(R - B/\Delta p),$$

- We get that only firms with capital at or above $\overline{A}(\gamma)$ can invest using direct finance.
- This is the classic **credit rationing** result going back to Stiglitz and Weiss (1981). The firm cannot get unlimited amounts of capital, for proper incentives to develop, it needs to have “skin in the game”.

Indirect Finance

- An intermediary can help relax the financing constraint of the firm by monitoring it and reducing its temptation to take the bad project.
- Now, the intermediary will get a share R_m of the return of the successful project

$$R_f + R_u + R_m = R$$

- The incentive constraint of the firm is now:

$$R_f \geq b/\Delta p$$

- There is also an incentive constraint for the intermediary:

$$R_m \geq c/\Delta p$$

- Then, the pledgeable expected income becomes:

$$p_H(R - (b + c)/\Delta p)$$

- Suppose that the intermediary is making a return of β (which has to exceed γ due to the monitoring cost), and invests I_m: $\beta = p_H R_m / I_m$, because of the incentive constraint it will contribute a least: $I_m(\beta) = p_H c/(\Delta p)\beta$.

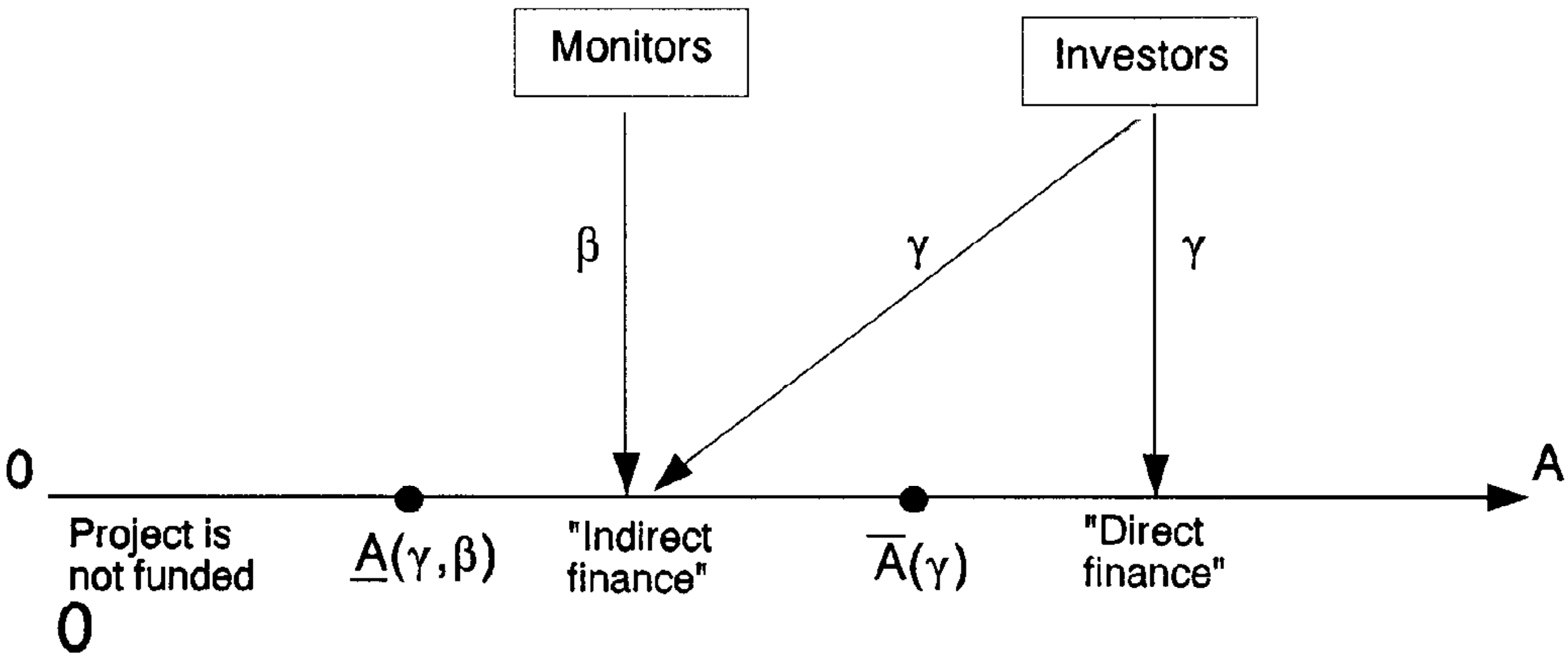
- Now, we can look at the financing constraint imposed by the participation constraint of the investors:

$$\gamma\big(I - A - I_m(\beta)\big) \leq p_H(R - (b + c)/\Delta p)$$

- This can be rewritten as:

$$A \geq \underline{A}(\gamma, \beta) = I - I_m(\beta) - p_H/\gamma(R - (b + c)/\Delta p)$$

- A firm with capital less than $\underline{A}(\gamma, \beta)$ cannot convince investors to supply it with capital even in the presence of intermediation. The firm will not increase reliance on intermediaries as their capital is more expensive.



- There are conditions in the paper guaranteeing that $\underline{A}(\gamma, \beta)$ is below $\overline{A}(\gamma)$.
- The result is that small firms are not financed at all, intermediate firms are financed by intermediaries and investors, and large firms are finance solely by investors.
- In equilibrium, the demand for capital equals the supply.
- The authors analyze the effects of decrease in the supply of capital.
- The main result is that the small firms are hurt most, as the squeeze leads to an increase in $\underline{A}(\gamma, \beta)$.