depositors (or borrowers) will tend to form coalitions and will buy (or sell) together in order to divide the transaction costs. (This argument does not work with proportional transaction costs.) Similarly, because of indivisibilities, a coalition of investors will be able to hold a more diversified (and thus less risky) portfolio than the ones individual investors would hold on their own.

Another type of scale economy is related to liquidity insurance à la Diamond and Dybvig (see Section 2.2 and Chapter 7). By the law of large numbers, a large coalition of investors will be able to invest in illiquid but more profitable securities, while preserving enough liquidity to satisfy the needs of individual investors. Once more this argument is not specific to the banking industry: it is also valid for insurance activities and more generally for inventory management. To have a genuine specificity of banks (as opposed to other intermediaries) informational asymmetries must again be introduced. This will be done in Section 2.3 in the discussion of the signaling approach, originally advanced by Leland and Pyle (1977). These informational asymmetries are also crucial for explaining the superiority of banks over financial markets in the provision of liquidity insurance.

2.2 Liquidity Insurance

A very natural idea for justifying the existence of depository institutions is to consider them as "pools of liquidity" that provide households with insurance against idiosyncratic shocks that affect their consumption needs. As long as these shocks are not perfectly correlated, the total cash reserve needed by a bank of size N (interpreted as a coalition of N depositors) increases less than proportionally with N. This is the basis for the "fractional reserve system," in which some fraction of the deposits can be used to finance profitable but illiquid investments. However, this is also the source of a potential fragility of banks, in the event that a high number of depositors decide to withdraw their funds for reasons other than liquidity needs. An interesting modeling of these issues, put forth by Diamond and Dybvig (1983), will be presented in detail in Chapter 7. For the moment, a simplified version of this model will be presented in order to capture the notion of liquidity insurance that was initially modeled by Bryant (1980).

2.2.1 The Model

Consider a one-good, three-period economy in which a continuum of ex-ante identical agents is each endowed with one unit of good at period $t = 0$, and this good is to be consumed at periods $t = 1$ and $t = 2$. The simplest way to model "liquidity shocks" is to consider that consumers learn at $t = 1$ whether they will have to consume *early* (i.e., at $t = 1$), in which case their utility function

is $u(C_1)$, or *late* (at $t = 2$), in which case their utility function is $\rho u(C_2)$ (where $\rho < 1$ is a discount factor). In ex-ante terms the expected utility of a depositor is

$$U = \pi_1 u(C_1^1) + \pi_2 \rho u(C_2^2), \tag{2.1}$$

where π_1 (resp. π_2) is the probability of being of "type 1" (resp. type 2) that is having to consume early (resp.late), and C_t^i denotes the consumption of an agent of type i at date t.[9]

u is assumed to be increasing and concave. Notice that preferences are state contingent and do not fit the standard Von Neumann–Morgenstern representation (they would if ρ were equal to one).

The good can be stored from one period to the next or can be invested at $t = 0$ in a long-run technology, which returns $R > 1$ units at $t = 2$, but only $L < 1$ units if it has to be liquidated at $t = 1$. The following discussion will compare different institutional arrangements and will show that a depository institution can improve the efficiency of the economy.

2.2.2 Autarky

The simplest case, in which there is no trade between agents, is called "autarky." Each agent chooses independently the quantity I that will be invested in the illiquid technology, assumed to be perfectly divisible. If he has to consume early, then this investment will be liquidated at $t = 1$, yielding

$$C_1 = 1 - I + LI = 1 - I(1 - L) \leq 1, \tag{2.2}$$

with equality only when $I = 0$. On the contrary, if he has to consume late, he obtains

$$C_2 = 1 - I + RI = 1 + I(R - 1) \leq R, \tag{2.3}$$

with equality only when $I = 1$.

In autarky, each consumer will select the consumption profile that maximizes his ex-ante utility U (given by 2.1) under the constraints 2.2 and 2.3.

2.2.3 Market Economy

If agents are allowed to trade, welfare improves. In this simple context, it is enough to open at $t = 1$ a financial market (for, say, a bond) in which agents can trade the good at $t = 1$ against a riskless bond (that is, a promise to receive some quantity of the consumption good at $t = 2$). Let p denote the price at $t = 1$ of the bond which, by convention, yields one unit of good at $t = 2$. Clearly $p \leq 1$; otherwise people would prefer to store. By investing I at $t = 0$, an agent can now obtain

$$C_1 = 1 - I + pRI, \tag{2.4}$$

if she needs to consume early (in which case she will sell RI bonds). If, on the contrary, she needs to consume late, she will obtain

$$C_2 = \frac{1-I}{p} + RI = \frac{1}{p}[1 - I + pRI], \tag{2.5}$$

since she can then buy $\frac{1-I}{p}$ bonds at $t = 1$. Since I can be freely chosen by agents, the only possible equilibrium price is $p = \frac{1}{R}$. Otherwise either an excess supply or an excess demand of bonds will occur ($I = +\infty$ if $p > \frac{1}{R}$, $I = 0$ if $p < \frac{1}{R}$). The equilibrium allocation of the market economy is therefore $C_1^M = 1$, $C_2^M = R$ and the corresponding investment level is $I^M = \pi_2$. Notice that this market allocation Pareto dominates the autarky allocation (see 2.2 and 2.3) since there is no liquidation. However as the next subsection will show, it is not ex-ante Pareto optimal.

2.2.4 Optimal Allocation

From an ex-ante viewpoint, there is a unique symmetric Pareto optimal allocation (C_1^*, C_2^*) obtained by solving

$$\max \pi_1 u(C_1) + \rho\pi_2 u(C_2) \tag{2.6}$$

$$\pi_1 C_1 + \pi_2 \frac{C_2}{R} = 1. \tag{2.7}$$

This optimal allocation satisfies in particular the first-order condition:

$$u'(C_1^*) = \rho R u'(C_2^*). \tag{2.8}$$

Therefore, except in the very peculiar case in which

$$u'(1) = \rho R u'(R),$$

the market allocation ($C_1^M = 1, C_2^M = R$) is not Pareto optimal. In particular, Diamond and Dybvig (1983) assume that $C \to Cu'(C)$ is decreasing.[10] In that case, since $R > 1$,

$$\rho R u'(R) < \rho u'(1) < u'(1), \tag{2.9}$$

and the market allocation can be Pareto improved by increasing C_1^M and decreasing C_2^M:

$$C_1^M = 1 < C_1^*; \; C_2^M = R > C_2^*. \tag{2.10}$$

In other words, the market economy does not provide perfect insurance against liquidity shocks, and therefore does not lead to an efficient allocation of resources. The following discussion will show how a financial intermediary can solve this problem.

2.2.5 Financial Intermediation

Provided the possibility of strategic behavior of depositors is ruled out (this issue will be studied in Chapter 7), the Pareto optimal allocation (C_1^*, C_2^*) can be implemented very easily by a financial intermediary who offers a demand deposit contract stipulated as follows: in exchange for a deposit of one unit at $t = 0$, individuals can get either C_1^* at $t = 1$ or C_2^* at $t = 2$. In order to fulfill its obligations, the FI stores $\pi_1 C_1^*$ and invests the rest in the illiquid technology. Thus we have established the following:

Result 2.1 In an economy in which agents are individually subject to independent liquidity shocks, the market allocation can be improved by a deposit contract offered by a financial intermediary.

The reason why the market allocation is not Pareto optimal is that complete contingent markets cannot exist: the state of the economy (i.e., the complete list of the consumers who need to consume early) is not observable by anyone. The only (noncontingent) financial market that can be opened (namely the bond market) is not sufficient to obtain efficient risk sharing.

Notice that a crucial assumption is that no individual withdraws at $t = 1$ if he or she does not have to. Provided $\rho R > 1$, this assumption is not unreasonable, since it corresponds to a Nash equilibrium behavior. Indeed 2.8 implies (since $\rho R > 1$) that $C_1^* < C_2^*$: in other words, a deviation by a single late consumer (withdraw at $t = 1$ and store the good until $t = 2$) is never in that consumer's own interest. However, Chapter 7 will show that another Pareto-dominated Nash equilibrium exists in which deviations of all late consumers occur simultaneously. Notice also that an FI cannot coexist (in this simple setup) with a financial market. Indeed, if there is a bond market at $t = 1$, the equilibrium price is necessarily $p = \frac{1}{R}$. Then the optimal allocation (C_1^*, C_2^*) is not a Nash equilibrium anymore: indeed 2.10 implies that

$$RC_1^* > R > C_2^*,$$

which means that late consumers are better off withdrawing early and buying bonds. This is of course a serious weakness of the model. Von Thadden (1994, 1996, 1997) has studied this question in a more general formulation that will be discussed in Chapter 7.

2.3 Information Sharing Coalitions

The common assumption for all the models presented in this section is that entrepreneurs are better informed than investors about the "quality" of the projects they want to develop. This "hidden information," or "adverse selection," paradigm will be explored in detail in several chapters of this book. The current discussion will show that this adverse selection paradigm can generate scale economies in the borrowing–lending activity, allowing interpretation of FIs as information sharing coalitions. After introducing (in subsection 2.3.1) a basic model of capital markets with adverse selection (that will be repeatedly used, under many variants, in several sections of this book), the seminal contribution of Leland and Pyle (1977) will be discussed in subsection 2.3.2. Leland and Pyle consider that entrepreneurs can "signal" the quality of their projects by investing more or less of their own wealth into these projects. In this way, they can partially overcome the adverse selection problem, since "good" projects can be separated from "bad" projects by their level of self-financing. However, if entrepreneurs are risk averse, this "signaling" is costly, since "good" entrepreneurs are obliged to retain a substantial fraction of the risk of their project. Leland and Pyle then study coalitions of borrowers and show that the "signaling cost" increases less rapidly than the size of the coalition. In other words, if borrowers form "partnerships," which Leland and Pyle interpret as FIs, they are able to obtain better financing conditions than by borrowing individually. This property is explained in subsection 2.3.3, and then several related contributions are summarized in subsection 2.3.4.

2.3.1 A Basic Model of Capital Markets with Adverse Selection

The following model of competitive capital markets with adverse selection will be used in several sections of this book. Consider a large number of entrepreneurs, each endowed with a risky project, requiring a fixed investment of a size normalized to one. The net returns $\tilde{R}(\theta)$ of these investments follow a normal distribution of mean θ and variance σ^2. Whereas σ^2 is the same for all projects, θ differs across projects and is the private information of each entrepreneur. However, the statistical distribution of θ in the population of entrepreneurs is common knowledge. The investors are risk neutral and have access to a costless storage technology. The entrepreneurs have enough initial wealth W_0 to finance their projects ($W_0 > 1$), but they would prefer to sell these projects because they are risk adverse. They have an exponential Von Neumann–Morgenstern utility function $u(w) = -e^{-\rho w}$, where w denotes their final wealth and $\rho > 0$ is their (constant) absolute index of risk aversion. If θ were observable, each entrepreneur would sell its project to the market at a price $P(\theta) = E[\tilde{R}(\theta)] = \theta^{11}$ and would be perfectly insured.[12] The final wealth of an entrepreneur of type θ would be $W_0 + \theta$.

Suppose now that θ is private information and that entrepreneurs are indistinguishable by investors. As in Akerlof (1970), the price P of equity will be the same for all firms, and in general only entrepreneurs with a lower expected return will sell their project. Indeed, by self-financing its project, entrepreneur θ obtains[13]

$$Eu(W_0 + \tilde{R}(\theta)) = u(W_0 + \theta - \frac{1}{2}\rho\sigma^2),$$

whereas by selling it to the market, he obtains $u(W_0 + P)$. Therefore entrepreneur θ will go to the financial market if and only if

$$\theta < \hat{\theta} = P + \frac{1}{2}\rho\sigma^2. \tag{2.11}$$

This means that only those entrepreneurs with a relatively low expected return ($\theta < \hat{\theta}$) will issue equity: this is exactly the adverse selection problem.

At equilibrium, the average return on equity will be equal to P (because of the investors' risk neutrality):

$$P = E[\theta | \theta < \hat{\theta}]. \tag{2.12}$$

The equilibrium of the capital market with adverse selection is thus characterized by a price of equity P and a cutoff level $\hat{\theta}$ such that relations 2.11 and 2.12 are satisfied. In general, the equilibrium outcome is inefficient. Assume, for instance, that the distribution of θ is binomial.[14] In other words, θ can take only two values: a low value θ_1 with probability π_1, and a high value θ_2 with probability π_2. Since the investors are risk neutral and the entrepreneurs are risk averse, efficiency requires that all entrepreneurs obtain 100 percent outside finance. By definition of the cutoff level, this means that $\hat{\theta} \geq \theta_2$. In that case, the price of equity equals

$$P = E[\theta] = \pi_1\theta_1 + \pi_2\theta_2.$$

Using 2.11 we obtain that this is only possible when

$$\pi_1\theta_1 + \pi_2\theta_2 + \frac{1}{2}\rho\sigma^2 \geq \theta_2,$$

or

$$\pi_1(\theta_2 - \theta_1) \leq \frac{1}{2}\rho\sigma^2. \tag{2.13}$$

In other words, the risk premium has to outweigh the adverse selection effect. If 2.13 is not satisfied, some entrepreneurs will prefer to self-finance, and the equilibrium outcome will be inefficient.[15]

2.3.2 Signaling Through Self-Financing

When 2.13 is not satisfied, the entrepreneurs who are endowed with good-quality projects ($\theta = \theta_2$) prefer to self-finance rather than to sell the entirety of their projects at a low price $P = E[\theta]$. In fact, they can limit themselves with partial self-finance if they can convince investors that the other entrepreneurs (who are endowed with low-quality projects, $\theta = \theta_1$) have no interest in doing the same (to "mimic" them, in the terminology of adverse selection models). In other words, deciding to self-finance a fraction α of the project will in that case "signal" to outside investors that this project is good. Intuitively, this is true when α is large enough. The "no mimicking" condition is:

$$u(W_0 + \theta_1) \geq Eu(W_0 + (1-\alpha)\theta_2 + \alpha\tilde{R}(\theta_1)). \tag{2.14}$$

The left side of 2.14 is the utility of a type θ_1 entrepreneur who sells all his project at a low price $P_1 = \theta_1$. The right side represents his expected utility when he mimics type θ_2, that is, sells only a fraction $(1-\alpha)$ of his project at a high unit price $P_2 = \theta_2$, but retains the risk on the remaining fraction α. With this model's assumption this expected utility equals $u(W_0 + (1-\alpha)\theta_2 + \alpha\theta_1 - \frac{1}{2}\rho\sigma^2\alpha^2)$, which gives a simplified version of 2.14:

$$\theta_1 \geq (1-\alpha)\theta_2 + \alpha\theta_1 - \frac{1}{2}\rho\sigma^2\alpha^2,$$

or

$$\frac{\alpha^2}{1-\alpha} \geq \frac{2(\theta_2 - \theta_1)}{\rho\sigma^2}. \tag{2.15}$$

Result 2.2 (Leland and Pyle 1977) When the level of projects' self-financing is observable, there is a continuum of signaling equilibria, parameterized by a number α fulfilling 2.15, and characterized by a low price of equity $P_1 = \theta_1$ for entrepreneurs who do not self-finance and a high price of equity $P_2 = \theta_2$ for entrepreneurs who self-finance a fraction α of their projects.

As is usual in signaling models (see Spence 1973), there is a continuum of equilibria, parameterized for the level α of self-financing by good-quality entrepreneurs. These equilibria can be Pareto ranked, since all lenders break even and θ_1 entrepreneurs get the same outcome as in the full-information case. As for θ_2 entrepreneurs, they obtain a utility level of $u(W_0 + \theta_2 - \frac{1}{2}\rho\sigma^2\alpha^2)$, instead of $u(W_0 + \theta_2)$ in the full-information case. Expressed in terms of lost income, their "informational cost of capital" is therefore

$$C = \frac{1}{2}\rho\sigma^2\alpha^2, \tag{2.16}$$

which is increasing in the level α of self-financing. The Pareto-dominating signaling equilibrium corresponds to the minimum possible value of α, which is defined implicitly by transforming 2.15 into an equality:

$$\frac{\alpha^2}{1-\alpha} = \frac{2(\theta_2 - \theta_1)}{\rho\sigma^2}. \tag{2.17}$$

It is natural to focus on this Pareto-dominating equilibrium, which allows definition of the (minimum) cost of capital

$$C(\sigma) = \frac{1}{2}\rho\sigma^2\alpha^2(\sigma) \tag{2.18}$$

where $\alpha(\sigma)$ is defined implicitly by 2.17.

2.3.3 Coalitions of Borrowers

This subsection will illustrate the main idea of this section, namely that in the presence of adverse selection, coalitions of borrowers can do better than individual borrowers. Suppose that N identical entrepreneurs of type θ_2 form a partnership and collectively issue securities in order to finance their N projects. If the individual returns of each project are independently distributed, and if the N entrepreneurs share equally both the proceeds of security issuing and the final returns, the situation is formally the same as before: the expected return per project is still θ_2, but (and this is the only difference) the variance per project is now $\frac{\sigma^2}{N}$ (because of diversification). Since one can prove that the function $\sigma \to C(\sigma)$ defined by 2.18 is increasing, the following result is obtained:

Result 2.3 (Diamond 1984) In the Leland-Pyle model (1977), the unit cost of capital decreases with the size of the coalition of borrowers (partnership or intermediary).

Proof It is necessary only to prove that $\sigma \to C(\sigma)$ is increasing. But relation 2.17 implies both that $\sigma \to \alpha(\sigma)$ is decreasing (since $\alpha \to \frac{\alpha^2}{1-\alpha}$ is increasing on [0, 1]) and that

$$\frac{1}{2}\rho\sigma^2\alpha^2(\sigma) = (\theta_2 - \theta_1)(1 - \alpha(\sigma)).$$

Therefore, 2.18 shows that $\sigma \to C(\sigma)$ is increasing, which was to be proved. ■

2.3.4 Related Justifications of FIs with Asymmetric Information

The Leland and Pyle (1977) model justifies FIs by considering the benefits obtained by borrowers when they form coalitions, provided they are able to communicate truthfully the quality of their projects within the coalition. But the framework of adverse selection (where the quality of projects is observable only by some investors) is sufficiently rich to study other possible justifications of FIs by coalition formation.

An agent endowed with private information faces two types of problems in order to benefit from this information. First, if she tries to sell her information directly she will be confronted with a classic credibility problem: the potential buyers may not be convinced that the information is true. Second, the profits she might obtain through trading on her information might be too small with respect to the cost of obtaining this information. These profits might even be zero