CHAPTER 1

# On the Role of Financial Markets and Institutions

Chapter Outline



## 1.1 Finance: The Time Dimension

Why do we need financial markets and institutions? We choose to address this question as our introduction to this text on financial theory. In doing so, we touch on some of the most difficult issues in finance and introduce concepts that will eventually require extensive development. Our purpose here is to phrase this question as an appropriate background for the study of the more technical issues that will occupy us at length. We also want to introduce some important elements of the necessary terminology. We ask the reader's patience as most of the sometimes difficult material introduced here will be taken up in more detail in the following chapters.

Fundamentally, a financial system is a set of institutions and markets permitting the exchange of contracts and the provision of services for the purpose of allowing the income and consumption streams of economic agents to be desynchronized—i.e., made less similar. It can, in fact, be argued that indeed the *primary* function of the financial system is to permit such desynchronization. There are two dimensions to this function: the time dimension and the risk dimension. Let us start with time. Why is it useful to disassociate consumption and income across time? Two reasons come immediately to mind. First, and somewhat trivially, income is typically received at discrete dates, say monthly, while it is customary to wish to consume continuously (i.e., every day).

Second, and more importantly, consumption spending defines a *standard of living*, and most individuals find it difficult to alter their standard of living from month to month or even from year to year. There is a general, if not universal, desire for a *smooth* consumption stream. Because it deeply affects everyone, the most important manifestation of this desire is the need to save (consumption smaller than income) for retirement so as to permit a consumption stream in excess of income (dissaving) after retirement begins. The *life-cycle* patterns of income generation and consumption spending are not identical, and the latter must be created from the former. The same considerations apply to shorter horizons. Seasonal patterns of consumption and income, for example, need not be identical. Certain individuals (car salespersons, department store salespersons, construction workers) may experience variations in income arising from seasonal events (e.g., most new cars are purchased in the spring and summer; construction activity is much reduced in winter), which they do not like to see transmitted to their ability to consume. There is also the problem created by temporary layoffs due to variation in aggregate economic activity that we refer to as business cycle fluctuations. While they are temporarily laid off and without substantial income, workers do not want their family's consumption to be severely reduced (Box 1.1).

Furthermore, and this is quite crucial for the growth process, some people—entrepreneurs, in particular—are willing to accept a relatively small income (but not necessarily

**BOX 1.1 Representing Preference for Smoothness**

The preference for a smooth consumption stream has a natural counterpart in the form of the utility function, $U(\ )$, which is typically used to represent the relative benefit a consumer receives from a specific consumption bundle. Suppose the representative individual consumes a single consumption good (or a basket of goods) in each of two periods, now and tomorrow. Let $c_1$ denote today's consumption level and $c_2$ tomorrow's, and let $U(c_1) + U(c_2)$ represent the level of utility (benefit) obtained from a given consumption stream $(c_1, c_2)$.

Preference for consumption smoothness must mean, for instance, that the consumption stream $(c_1,c_2) = (4,4)$ is preferred to the alternative $(c_1, c_2) = (3, 5)$, or

$$U(4) + U(4) > U(3) + U(5)$$

Dividing both sides of the inequality by 2, this implies

$$U(4) > \frac{1}{2}U(3) + \frac{1}{2}U(5)$$

(*Continued*)

**BOX 1.1 Representing Preference for Smoothness (Continued)**

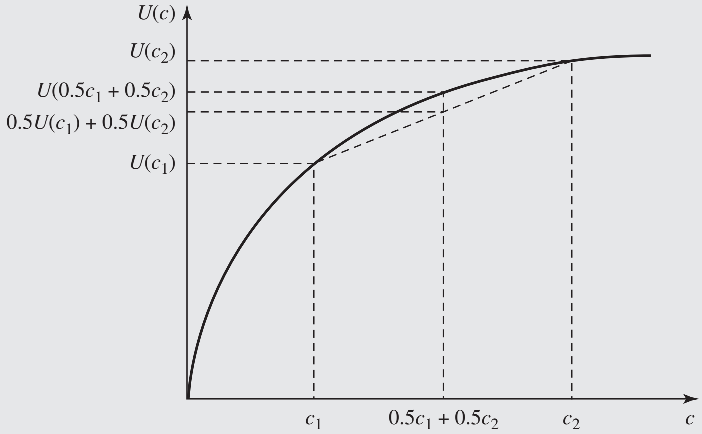


**Figure 1.1**
A strictly concave utility representation.

As shown in Figure 1.1, when generalized to all possible alternative consumption pairs, this property implies that the function $U(\cdot)$ has the rounded shape that we associate with the term *strict concavity*.

consumption!) for an initial period of time in exchange for the prospect of high returns (and presumably high income) in the future. They are operating a sort of arbitrage over time. This does not disprove their desire for smooth consumption; rather, they see opportunities that lead them to accept what is formally a low-income level initially against the prospect of a much higher income level later (followed by a zero income level when they retire). They are investors who, typically, do not have enough liquid assets to finance their projects and, as a result, need to raise capital by borrowing or by selling shares.

Indeed, the first key element in finance is **time.** In a timeless world, there would be no assets, no financial transactions (although money would be used, it would have only a transaction function), and no financial markets or institutions. The very notion of a security (a financial contract) implies a time dimension.

Asset holding permits the desynchronization of consumption and income streams. The peasant putting aside seeds, the miser burying his gold, or the grandmother putting a few hundred dollar bills under her mattress are all desynchronizing their consumption and income, and in doing so, presumably seeking a higher level of well-being for themselves. A fully developed financial system should also have the property of fulfilling this same function *efficiently*. By that we mean that the financial system should provide versatile and diverse instruments to accommodate the widely differing needs of savers and borrowers insofar as size (many small lenders, a few big borrowers), timing, and maturity of loans (how to finance long-term projects with short-term money), and the liquidity characteristics of instruments (precautionary saving cannot be tied up permanently). In other words, the elements composing the financial system should aim at *matching* the diverse financing needs of different economic agents as perfectly as possible.

## 1.2 Desynchronization: The Risk Dimension

We have argued that time is of the essence in finance. When we talk of the importance of time in economic decisions, we think in particular of the relevance of choices involving the present versus the future. But the future is, by its very nature, uncertain: financial decisions with implications (payouts) in the future are necessarily risky. Time and risk are inseparable. This is why **risk** is the second key word in finance.

For the moment, let us compress the time dimension into the setting of a "Now and Then" (present versus future) economy. The typical individual is motivated by the desire to smooth consumption between "Now" and "Then." This implies a desire to identify consumption opportunities that are as similar as possible among the different possibilities that may arise "Then." In other words, *ceteris paribus*—most individuals would like to guarantee their family the same standard of living whatever events transpire tomorrow: whether they are sick or healthy, unemployed or working, confronted with bright or poor investment opportunities, fortunate or hit by unfavorable accidental events.[1] This characteristic of preferences is generally described as "aversion to risk."

A productive way to start thinking about this issue is to introduce the notion of *states of nature* or *states of the world*. A state of nature is a complete description of a possible scenario for the future across all the dimensions relevant for the problem at hand. In a "Now and Then" economy, all possible future events can be represented by an exhaustive list of states of nature. We can thus extend our former argument for smoothing consumption across time by noting that the typical "risk averse" individual would also like to experience similar consumption levels across all future states of nature, whether good or bad.

An efficient financial system offers ways for savers to reduce or eliminate, at a fair price, the risks they are not willing to bear (risk shifting). Fire insurance contracts eliminate the financial risk of fire, while put options contracts can prevent the loss in wealth associated with a stock's price declining below a predetermined level, to mention but two examples. The financial system also makes it possible to obtain relatively safe aggregate returns from a large number of small, relatively risky investments. This is the process of diversification. By permitting economic agents to *diversify*, to *insure*, and to *hedge* their risks, an efficient

[1] *Ceteris paribus* is the Latin phrase for "everything else maintained equal." It is an expression commonplace in the language of economics.

financial system fulfills the function of redistributing purchasing power not only over time, but also across states of nature.[2]

## 1.3 The Screening and Monitoring Functions of the Financial System

The business of desynchronizing consumption from income streams across time and states of nature is often more complex than our initial description may suggest. If time implies uncertainty, uncertainty may imply not only risk, but often *asymmetric information* as well. By this term, we mean situations where the agents involved have different information, with some being potentially better informed than others. How can a saver be assured that he will be able to find a borrower with a good ability to repay—the borrower himself knows more about this, but he may not wish to reveal all he knows—or an investor (an entrepreneur or a firm) with a good project, yielding the most attractive return for him and hopefully for society as well? Again, the investor is likely to have a better understanding of the project's prospects and of his own motivation to carry it through. What do "good" and "most attractive" mean in these circumstances? Do these terms refer to the highest potential return? What about risk? What if the anticipated return is itself affected by the actions of the investors themselves (a phenomenon labeled "moral hazard")? How does one share the risks of a project in such a way that both investors and savers are willing to proceed, taking actions acceptable to both? It is the task of financial intermediaries—banks, venture capital firms, and private equity firms—to answer these questions and to do so in such a way that brings the socially beneficial projects to fruition. An efficient financial system, and the financial institutions that define it, not only assists in these information and monitoring tasks, but also provides a range of instruments (contractual arrangements) suitable for the largest number of savers and borrowers, thereby contributing to the channeling of savings toward the most efficient projects.[3]

In the words of the preeminent economist, Joseph Schumpeter (1934), "Bankers are the gatekeepers of capitalist economic development. Their strategic function is to screen potential innovators and advance the necessary purchasing power to the most promising."

[2] Both insurance and hedging are risk-reduction strategies but with one critical difference. In the case of insurance, the investor pays money—the insurance premium—to guarantee against a loss in value of some asset that he owns (a house, shares of stock). In the case of hedging, an investor adds to his portfolio, usually at very little cost, another asset (the "hedging asset") with a price pattern that is opposite to that of his original portfolio: if the original portfolio declines in value, the newly added asset increases in value by an equal and offsetting amount (this is the case of a "perfect hedge"). The opposite is true, however, if the investor's original portfolio increases in value: the hedging asset loses an equal and offsetting amount. The investor thus sacrifices potential gains to his portfolio's value in exchange for protection against losses. In the case of insurance, upward potential was not sacrificed, but the investor had to pay the premium.

[3] If the extent of the information asymmetry between buyers and sellers in a market becomes too great, the market may shut down: no trades occur. This exact event occurred at the start of the financial crisis when the investment banks that had been packaging pools of US home mortgages into mortgage-backed securities (MBS) discovered that they could find no buyers. The natural buyers of these securities had become suspicious that their quality was not as advertised. The forced sale of a nearly insolvent Bear Stearns to JPMorgan Chase and the Lehman Brothers bankruptcy ensued.

For highly risky projects, such as the creation of a new firm exploiting a new technology, venture capitalists largely provide this function today.

## 1.4 The Financial System and Economic Growth

The performance of the financial system matters at several levels. We shall argue that it matters for growth, that it impacts the characteristics of the business cycle, and, most importantly, that it is a significant determinant of economic welfare. We tackle growth first. Channeling funds from savers to investors efficiently is obviously important. Whenever more efficient ways are found to perform this task, society can achieve a greater increase in tomorrow's consumption for a given sacrifice in consumption today. As a result, savings becomes a more attractive alternative to current consumption, and households save more. Intuitively, more savings should lead to greater investment and thus greater future wealth. Figure 1.2 indeed suggests that, for 90 developing countries over the period 1971–1992, there was a strong positive association between saving rates and growth rates. When looked at more carefully, however, the evidence is usually not as strong.[4]

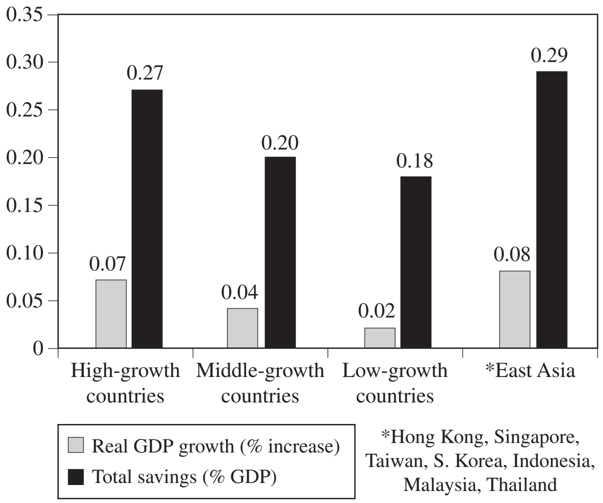


**Figure 1.2**
Savings and growth in 90 developing countries.

[4] In a straightforward regression in which the dependent variable is the growth rate in real per capita gross national product (GNP), the coefficient on the average fraction of real GNP represented by investment ($I/Y$) over the prior 5 years is positive but insignificant. Together with other results, this is interpreted as suggesting a possible reverse causation from real per capita GNP growth to investment spending. See Barro and Sala-i-Martin (1995), Chapter 12, for a full discussion. There is also a theoretically important distinction between the effects of increasing investment (savings) (as a proportion of national income) on an economy's *level* of wealth and its *growth rate*. Countries that save more will ceteris paribus be wealthier, but they need not grow more rapidly. The classic growth model of Solow (1956) illustrates this distinction. See Web Chapter A.

One important reason may be that the hypothesized link is, of course, dependent on a ceteris paribus clause: It applies only to the extent savings are invested in appropriate ways. The economic performance of the former USSR reminds us that it is not enough only to save; it is also important to invest judiciously. Historically, the investment/GDP (gross domestic product, the measure of a nation's aggregate economic output) ratio in the USSR was very high in international comparisons, suggesting the potential for very high growth rates.[5] After 1989, however, experts realized that the value of the existing stock of capital was not consistent with the former levels of investment. A great deal of the investment had been effectively wasted—in other words, allocated to poor or even worthless projects. Equal savings rates can thus lead to investments of widely differing degrees of usefulness from the viewpoint of future growth.

A more contemporary version of this same overinvestment phenomenon may be present in China. In 2012, investment in China represented 46% of output which is astonishingly high by international standards (e.g., the United States averages 15%; Switzerland averages around 20%), and a lively debate has arisen as to what fraction of China's investment will ultimately be useful.[6]

Let us go further than these general statements in the analysis of the savings and growth nexus and of the role of the financial system. Following Barro and Sala-i-Martin (1995), one can view the process of transferring funds from savers to investors in the following way.[7] The least efficient system would be one in which all investments are made by the savers themselves. This is certainly inefficient because it requires a sort of "double coincidence" of intentions: good investment ideas occurring in the mind of someone lacking past savings will not be realized. Funds that a nonentrepreneur saves would not be put to productive use. Yet, this unfortunate situation is a clear possibility if the necessary confidence in the financial system is lacking, with the consequence that savers do not entrust the system with their savings. One can thus think of circumstances where savings never enter the financial system, or where only a small fraction does. When it does, it will typically enter via some sort of depository institution. In an international setting, a similar problem arises if national savings are primarily invested abroad, a situation that may reach alarming proportions in the case of underdeveloped

5 More precisely, GDP refers to the value, at market prices, of all final goods and services (those sold to end users) produced within a nation's geographical boundaries during a specific time period (usually a year).

6 Excessive investment comes at the price of lower household consumption. Lee et al. (2012) estimate this loss in consumption to have averaged 4% of total Chinese output.

7 For a broader perspective and a more systematic connection with the relevant literature on this topic, see Levine (1997).

countries.[8] Let FS/$S$ represent, then, the fraction of aggregate savings ($S$) being entrusted to the financial system (FS).

At a second level, the functioning of the financial system may be more or less costly. While funds transferred from a saver to a borrower via a direct loan are immediately and fully made available to the end user, the different functions of the financial system discussed above are often best fulfilled, or sometimes can only be fulfilled, through some form of intermediation, which typically involves some cost. Let us think of these costs as administrative costs, on the one hand, and costs linked to the reserve requirements of banks, on the other. Different systems will have different operating costs in this large sense, and, as a consequence, the amount of resources transferred to investors will also vary. Let us think of BOR/FS as the ratio of funds transferred from the financial system to borrowers and entrepreneurs.

Borrowers themselves may make diverse use of the funds borrowed. Some, for example, may have pure liquidity needs (analogous to the reserve needs of depository institutions), and if the borrower is the government, it may well be borrowing for consumption! For the savings and growth nexus, the issue is how much of the borrowed funds actually result in productive investments. Let $I$/BOR represent the fraction of borrowed funds actually invested. Note that BOR stands for borrowed funds whether private or public. In the latter case, a key issue is what fraction of the borrowed funds are used to finance public investment as opposed to public consumption.

Finally, let EFF denote the efficiency of the investment projects undertaken in society at a given time, with EFF normalized at unity; in other words, the average investment project has EFF = 1, the below-average project has EFF $<$ 1, and conversely for the above average project (a project consisting of building a bridge leading nowhere would have an EFF = 0); $K$ is the aggregate capital stock and $\Omega$ the depreciation rate. We may then write

$$\dot{K} = \mathrm{EFF} \cdot I - \Omega K \tag{1.1}$$

or, multiplying and dividing $I$ with each of the newly defined variables

$$\dot{K} = \mathrm{EFF} \cdot (I/\mathrm{BOR}) \cdot (\mathrm{BOR}/\mathrm{FS}) \cdot (\mathrm{FS}/S) \cdot (S/Y) \cdot Y - \Omega K \tag{1.2}$$

[8] The problem is slightly different here, however. Although capital flight is a problem from the viewpoint of building up a country's domestic capital stock, the acquisition of foreign assets may be a perfectly efficient way of building a national capital stock. The effect on growth may be negative when measured in terms of GDP, but not necessarily so in terms of national income or GNP. Switzerland is an example of a rich country investing heavily abroad and deriving a substantial income flow from it. It can be argued that the growth rate of the Swiss GNP (but probably not GDP) has been enhanced rather than diminished by this fact.

where our notation is meant to emphasize that the growth of the capital stock at a given savings rate is likely to be influenced by the levels of the various ratios introduced above.[9] Let us now review how this might be the case.

One can see that a financial system performing its matching function efficiently will positively affect the savings rate ($S/Y$) and the fraction of savings entrusted to financial institutions (FS/$S$). This reflects the fact that savers can find the right savings instruments for their needs. In terms of overall services net of inconvenience, this acts like an increase in the return to the fraction of savings finding its way into the financial system. The matching function is also relevant for the $I$/BOR ratio. With the appropriate instruments (like flexible overnight loan facilities), a firm's cash needs are reduced and a larger fraction of borrowed money can actually be used for investment.

By offering a large and diverse set of possibilities for spreading risks (insurance and hedging), an efficient financial system will also positively influence the savings ratio ($S/Y$) and the FS/$S$ ratio. Essentially this works through improved return/risk opportunities, corresponding to an improved trade-off between future and present consumption (for savings intermediated through the financial system). Furthermore, in permitting entrepreneurs with risky projects to eliminate unnecessary risks by using appropriate instruments, an efficient financial system provides, somewhat paradoxically, a better platform for undertaking riskier projects. If, on average, riskier projects are also the ones with the highest returns, as most of financial theory reviewed later in this book leads us to believe, one would expect that the more efficiently this function is performed, the higher (ceteris paribus) the value of EFF; in other words, the higher, on average, the efficiency of the investment undertaken with the funds made available by savers.

Finally, a more efficient system may be expected to screen alternative investment projects more effectively and to monitor more thoroughly and more cost efficiently the conduct of the investments (efforts of investors). The direct impact is to increase EFF. Indirectly this also means that, on average, the return/risk characteristics of the various instruments offered savers will be improved and one may expect, as a result, an increase in both $S/Y$ and FS/$S$ ratios.

The previous discussion thus tends to support the idea that the financial system plays an important role in permitting and promoting the growth of economies.[10] Yet growth is not an objective in itself. There is such a thing as excessive capital accumulation typically

9 $\dot{K} = \mathrm{d}K/dt$, i.e., the change in $K$ as a function of time.

10 There is statistical support asserting the beneficial consequences of financial development for economic growth at the country (King and Levine, 1993a, b), industry (Rajan and Zingales, 1998), and firm levels (Aghion et al., 2007).

funded in some way by financial repression directed at households. In the case of Italy, Jappelli and Pagano (1994) suggest that household borrowing constraints,[11] in general a source of inefficiency and the mark of a less than perfect financial system, may have led to more savings than desired in the 1980s.

The excessive investment rates evident in China (noted earlier) are partially funded by high household savings (the household savings rate in China is presently approximately 25%) in part driven by financial repression in the form of government-mandated low interest rates that banks are permitted to offer depositors and prohibitions on the ownership of certain types of securities (e.g., stocks or bonds issued by foreign-based firms or foreign governments). In the absence of a significant social safety net (no government sponsored pensions or health care), households are thus induced to save a lot, much of which gets channeled into the investments of state-owned enterprises or local governments. Financial "repression" takes a different form in India: there are few banks in rural areas. As a result, many rural households do not commit their savings to the financial system at all, but prefer to buy "gold." "Investment" in the form of gold purchases contributes nothing to India's growth prospects.

While these examples are purely illustrative, they underscore the necessity of adopting a broader and more satisfactory viewpoint and of more generally studying the impact of the financial system on social welfare. This is best done in the context of the theory of general equilibrium, a subject to which we next turn.

## 1.5 Financial Markets and Social Welfare

Let us next consider the role of financial markets in the allocation of resources and, consequently, their effects on social welfare. The perspective provided here places the process of financial innovation in the context of the theory of general economic equilibrium whose central concepts are closely associated with the *Ecole de Lausanne* and the names of Léon Walras and Vilfredo Pareto.

Our starting point is the first theorem of welfare economics, which defines the conditions under which the allocation of resources implied by the general equilibrium of a decentralized competitive economy is efficient or optimal in the Pareto sense.

First, let us define the terms involved. Assume a timeless economy where a large number of economic agents interact. There is an arbitrary number of goods and services, *n*. Consumers possess a certain quantity (possibly zero) of each of these *n* goods (in particular, they have

[11] By "borrowing constraints," we mean the limitations that the average individual or firm may experience in his or her ability to borrow, at current market rates, from financial institutions.

the ability to work a certain number of hours per period). They can sell some of these goods and buy others at prices quoted in markets. There are a large number of firms, each represented by a production function—i.e., a given ability (constrained by what is technologically feasible) to transform some of the available goods or services (inputs) into others (outputs)—for instance, combining labor and capital to produce consumption goods. Agents in this economy act selfishly: Individuals maximize their well-being (utility), and firms maximize their profits.

General equilibrium theory tells us that, thanks to the action of the price system, order will emerge out of this uncoordinated chaos, provided certain conditions are satisfied. In the main, these hypotheses (conditions) are as follows:

H1: *Complete markets.* There exists a market on which a price is established for each of the $n$ goods valued by consumers.
H2: *Perfect competition.* The number of consumers and firms (i.e., demanders and suppliers of each of the $n$ goods in each of the $n$ markets) is large enough so that no agent is in a position to influence (manipulate) market prices; i.e., all agents take prices as given.
H3: Consumers' preferences are convex.
H4: Firms' production sets are convex as well.

H3 and H4 are technical conditions with economic implications. Somewhat paradoxically, the convexity hypothesis for consumers' preferences approximately translates into strictly concave utility functions. In particular, H3 is satisfied (in substance) if consumers display risk aversion, an assumption crucial for understanding financial markets, and one that will be made throughout this text. As already noted (Box 1.2), risk aversion translates into strictly concave utility functions (see Chapter 4 for details). H4 imposes requirements on

**BOX 1.2 Representing Risk Aversion**

Let us reinterpret the two-date consumption stream $(c_1, c_2)$ of Box 1.1 as the consumption levels attained "Then" or "Tomorrow" in two alternative, equally likely, states of the world. The desire for a smooth consumption stream across the two states, which we associate with risk aversion, is obviously represented by the same inequality

$$U(4) > \frac{1}{2}U(3) + \frac{1}{2}U(5)$$

and it implies the same general shape for the utility function. In other words, assuming plausibly that decision makers are **risk averse**, an assumption in conformity with most of financial theory, implies that the utility functions used to represent agents' preferences are **strictly concave.**

the production technology. It specifically rules out increasing returns to scale in production. Although important, this assumption is nevertheless not at the heart of things in financial economics since for the most part we will abstract from the production side of the economy.

A **general competitive equilibrium** is a price vector $p^*$ and an allocation of resources, resulting from the independent decisions of consumers and producers to buy or sell each of the $n$ goods in each of the $n$ markets, such that, at the equilibrium price vector $p^*$, supply equals demand in all markets simultaneously and the action of each agent is the most favorable to him or her among all those he can afford (technologically or in terms of his budget computed at equilibrium prices).

A **Pareto optimum** is an allocation of resources, however determined, where it is impossible to redistribute resources (i.e., to go ahead with further exchanges) without reducing the welfare of at least one agent. In a Pareto-efficient (or Pareto optimal—we will use the two terminologies interchangeably) allocation of resources, it is thus not possible to make someone better off without making someone else worse off. Such a situation may not be just or fair, but it is certainly efficient in the sense of avoiding waste.

Omitting some purely technical conditions, the main results of general equilibrium theory can be summarized as follows:

1. *The existence of a competitive equilibrium*: Under H1 through H4, a competitive equilibrium is guaranteed to exist. This means that there indeed exists a price vector and an allocation of resources satisfying the definition of a competitive equilibrium as stated above.
2. *First welfare theorem*: Under H1 and H2, a competitive equilibrium, if it exists, is a Pareto optimum.
3. *Second welfare theorem*: Under H1 through H4, any Pareto-efficient allocation can be decentralized as a competitive equilibrium.

The second welfare theorem asserts that, for any arbitrary Pareto-efficient allocation, there is a price vector and a set of initial endowments such that this allocation can be achieved as a result of the free interaction of maximizing consumers and producers interacting in competitive markets. To achieve a specific Pareto optimal allocation, some redistribution mechanism will be needed to reshuffle initial resources. The availability of such a mechanism, functioning without distortion (and thus waste), is, however, very much in question. Hence the dilemma between equity and efficiency that faces all societies and their governments.

The necessity of H1 and H2 for the optimality of a competitive equilibrium provides a rationale for government intervention when these hypotheses are not naturally satisfied. The case for antitrust and other “pro-competition” policies is implicit in H2; the case for intervention in the presence of externalities or in the provision of public goods follows from H1, because these two situations are instances of missing markets.[12]

Note that so far there does not seem to be any role for financial markets in promoting an efficient allocation of resources. To restore that role, we must abandon the fiction of a timeless world, underscoring, once again, the fact that time is of the essence in finance! Introducing the time dimension does not diminish the usefulness of the general equilibrium apparatus presented above, provided the definition of a good is properly adjusted to take into account not only its intrinsic characteristics, but also the time period in which it is available. A cup of coffee available at date $t$ is different from a cup of coffee available at date $t + 1$, and, accordingly, it is traded on a different market and it commands a different price. Thus, if there are two dates, the number of goods in the economy goes from $n$ to $2n$.

It is easy to show, however, that not all commodities need be traded for future as well as current delivery. The existence of a spot and forward market *for one good only* (taken as the numeraire) is sufficient to implement all the desirable allocations, and, in particular, restore, under H1 and H2, the optimality of the competitive equilibrium. This result is contained in Arrow (1964). It provides a powerful economic rationale for the existence of credit markets, markets where money is traded for future delivery.

Now let us go one step further and introduce uncertainty, which we will represent conceptually as a partition of all the relevant future scenarios into separate *states of nature*. To review, a state of nature is an exhaustive description of one possible relevant configuration of future events. Using this concept, we can extend the applicability of the welfare theorems in a fashion similar to that used with time above, by defining goods according not only to the date but also to the state of nature at which they are (might be) available. This is the notion of contingent commodities. Under this construct, we imagine the market for ice cream decomposed into a series of markets: for ice cream today, ice cream tomorrow if it rains and the Dow Jones is at 10,000; if it rains and so on. Formally, this is a straightforward extension of the basic context: there are more goods, but this is not in itself restrictive[13] (Arrow, 1964; Debreu, 1959).

[12] Our model of equilibrium presumes that agents affect one another only through prices. If this is not the case, an economic externality is said to be present. These may involve either production or consumption. For example, there have been substantial negative externalities for fishermen associated with the construction of dams in the western United States: the catch of salmon has declined dramatically as these dams have reduced the ability of the fish to return to their spawning habitats. If the externality affects all consumers simultaneously, it is said to be a public good. The classic example is national defense. If any citizen is to consume a given level of national security, all citizens must be equally secure (and thus consume this public good at the same level). Both are instances of missing markets. Neither is there a market for national defense nor for rights to disturb salmon habitats.

[13] In this context $n$ can be as large as one needs without restriction.

The hypothesis that there exists a market for each and every good valued by consumers becomes, however, much more questionable with this extended definition of a typical good, as the example above suggests. On the one hand, the number of states of nature is, in principle, arbitrarily large and, on the other, one simply does not observe markets where commodities contingent on the realization of individual states of nature can routinely be traded. One can thus state that *if* markets are complete in the above sense, a competitive equilibrium is efficient, but the issue of completeness (H1) then takes center stage. Can Pareto optimality be obtained in a less formidable setup than one where there are complete contingent commodity markets? What does it mean to make markets "more complete?"

It was Arrow (1964), again, who took the first step toward answering these questions. Arrow generalized the result alluded to earlier and showed that it would be enough, in order to effect all desirable allocations, to have the opportunity to trade one good only across all states of nature. Such a good would again serve as the numeraire. The primitive security could thus be a claim promising $1.00 (i.e., one unit of the numeraire) at a future date, contingent on the realization of a particular state, and zero under all other circumstances. We shall have a lot to say about such *Arrow−Debreu securities* (henceforth A−D securities), which are also called *contingent claims*. Arrow asserted that if there is one such contingent claim corresponding to each and every one of the relevant future date/state configurations, hypothesis H1 could be considered satisfied, markets could be considered complete, and the welfare theorems would apply. Arrow's result implies a substantial decrease in the number of required markets.[14] However, for a complete contingent claim structure to be fully equivalent to a setup where agents could trade a complete set of contingent commodities, it must be the case that agents are assumed to know all future spot prices, contingent on the realization of all individual states of the world. Indeed, it is at these prices that they will be able to exchange the proceeds from their A−D securities for consumption goods. This hypothesis is akin to the hypothesis of rational expectations.[15]

A−D securities are a powerful conceptual tool and are studied in depth in Chapters 9 and 11. They are not, however, the instruments we observe being traded in actual markets. Why is this the case, and in what sense is what we do observe an adequate substitute? To answer these questions, we first allude to a result (derived later on) which states that there is no single way to make markets complete. In fact, potentially a large number of alternative financial structures may achieve the same goal, and the complete A−D securities structure is only one of them. For instance, we shall describe, in Chapter 11, a context in which one might think of achieving an essentially complete market structure with options or derivative securities. We shall make use of this fact for pricing alternative instruments using arbitrage

14 Example: 2 dates, 3 basic goods, 4 states of nature: complete commodity markets require 12 contingent commodity markets plus 3 spot markets versus 4 contingent claims and $2 \times 3$ spot markets in the Arrow setup.

15 For an elaboration on this topic, see Drèze (1971).

techniques. Thus, the failure to observe anything close to A−D securities being traded is not evidence against the possibility that markets are indeed complete.

In an attempt to match this discussion on the role played by financial markets with the type of markets we see in the real world, one can identify the different needs met by trading A−D securities in a complete markets world. In so doing, we shall conclude that, in reality, different needs are met trading alternative specialized financial instruments (which, as we shall later prove, will all appear as portfolios of A−D securities).

As we have already observed, the time dimension is crucial for finance, and, correspondingly, the need to exchange purchasing power across time is essential. It is met in reality through a variety of specific *noncontingent* instruments, which are promised future payments independent of specific states of nature, except those in which the issuer is unable to meet his obligations (bankruptcies). Personal loans, bank loans, money market and capital market instruments, social security, and pension claims are all assets fulfilling this basic need for redistributing purchasing power in the time dimension. In a complete market setup implemented through A−D securities, the needs met by these instruments would be satisfied by a certain configuration of positions in A−D securities. In reality, the specialized instruments mentioned above fulfill the demand for exchanging income through time.

One reason for the formidable nature of the complete markets requirement is that a state of nature, which is a complete description of the relevant future for a particular agent, includes some purely personal aspects of almost unlimited complexity. Certainly the future is different for you, in a relevant way, if you lose your job, or if your house burns, without these contingencies playing a very significant role for the population at large. In a pure A−D world, the description of the states of nature should take account of these *individual contingencies* viewed from the perspective of each and every market participant! In the real world, insurance contracts are the specific instruments that deal with the need for exchanging income across purely individual events or states. The markets for these contracts are part and parcel of the notion of complete financial markets. Although such a specialization makes sense, it is recognized as unlikely that the need to trade across individual contingencies will be fully met through insurance markets because of specific difficulties linked with the hidden quality of these contingencies (i.e., the inherent asymmetry in the information possessed by suppliers and demanders participating in these markets). The presence of these asymmetries strengthens our perception of the impracticality of relying exclusively on pure A−D securities to deal with personal contingencies.

Beyond time issues and personal contingencies, most other financial instruments not only imply the exchange of purchasing power through time, but are also more specifically contingent on the realization of particular events. The relevant events here, however, are defined on a collective basis rather than being based on individual contingencies; they are contingent on the realization of events affecting groups of individuals and observable by everyone. An example is the situation where a certain level of profits for a firm implies the payment of a certain dividend against the ownership of that firm's equity. Another is the payment of a certain sum of money associated with the ownership of an option or a financial futures. In the later cases, the contingencies (sets of states of nature) are dependent on the value of the underlying asset itself.

## 1.6 Financial Intermediation and the Business Cycle

Business cycles are the mark of all developed economies. According to much of current research, they are in part the result of external shocks with which these economies are repeatedly confronted. The depth and amplitude of these fluctuations, however, may well be affected by some characteristics of the financial system. This is at least the import of the recent literature on the financial accelerator. The mechanisms at work here are numerous, and we limit ourselves to giving the reader a flavor of the discussion.

The financial accelerator is manifest most straightforwardly in the context of monetary policy implementation. Suppose the monetary authority wishes to reduce the level of economic activity (inflation is feared) by raising real interest rates. The primary effect of such a move will be to increase firms' cost of capital and, as a result, to induce a decrease in investment spending as marginal projects are eliminated from consideration.

According to the financial accelerator theory, however, there may be further, substantial, secondary effects. In particular, the interest rate rise will reduce the value of firms' collateralizable assets. For some firms, this reduction may significantly diminish their access to credit, making them credit constrained. As a result, the fall in investment may exceed the direct impact of the higher cost of capital; tighter financial constraints may also affect input purchases or the financing of an adequate level of finished goods inventories. For all these reasons, the output and investment of credit-constrained firms will be more strongly affected by the action of the monetary authorities, and the economic downturn may be made correspondingly more severe. By this same mechanism, any economywide reduction in asset values may have the effect of reducing economic activity under the financial accelerator.

Which firms are most likely to be credit constrained? We would expect that small firms, those for which lenders have relatively little information about the long-term prospects, would be principally affected. These are the firms from which lenders demand high levels of collateral. Bernanke et al. (1996) provide empirical support for this assertion using US data from small manufacturing firms.

The financial accelerator has the power to make an economic downturn, of whatever origin, more severe. If the screening and monitoring functions of the financial system can be tailored more closely to individual firm needs, lenders will need to rely to a lesser extent on collateralized loan contracts. This would diminish the adverse consequences of the financial accelerator and perhaps the severity of business cycle downturns.

## 1.7 Financial Crises

A more radical version of the link between the financial markets and the business cycle is present in the experience of a financial crisis. It reflects the notion of a financial crisis as either a catalyst for, or the initial cause of, a severe and prolonged business cycle downturn.[16] The Great Depression of the early 1930s and the recent "Great Recession" of 2007–2009 are only the most dramatic cases in point. Both were associated with a large decline in output, a dramatic decline in investment, and a large increase in the number of persons unemployed.

Financial crises of this magnitude are typically preceded by the ending of a price "bubble" in some important asset type, with ensuing price declines in it and many other asset categories. (The "Great Recession" in the United States, in particular, was preceded by the ending of a residential real estate price bubble chiefly in the states of the Southwest, California, Texas, and Florida. Since these homes had been purchased with large mortgages and very little equity, the prospect of numerous defaults immediately arose.) With bank assets (mortgages or MBSs in the Great Recession case) declining in value, many banks face the very real possibility of insolvency, and some collapse (Lehman Brothers).[17] In either case, banks quickly become reluctant to extend any further risky loans for fear of making their own financial situations even worse if the loans cannot be repaid. As a consequence, a "credit crunch" ensues whereby many firms, and especially small and medium sized ones, find it essentially impossible to obtain loans to fund their investments and continuing operations. See Figure 1.3 for some sense of the drop in US lending activity in the first years of the Great Recession. Aggregate investment spending "dries up." Note that the resulting economic contraction parallels the one associated with the financial accelerator of Section 1.6, except that it is typically more immediate and more intense. In effect, the financial system ceases to perform the functions assigned to it by society as described in Section 1.4. The aforementioned drop in investment spending is often accompanied by a reduction in consumption spending as households react to their reduced wealth resulting from the decline in asset prices. Output contracts further.

[16] For a detailed historical analysis, see Reinhart and Rogoff (2009).

[17] Lehman Brothers was not technically a (commercial) bank because it was not permitted to take deposits. Neither did it have the right to receive loans from the US Federal Reserve, the lender of last resort. Lehman Brothers funded its assets (real estate, MBSs) with very short-term loans that had to be rolled over daily.

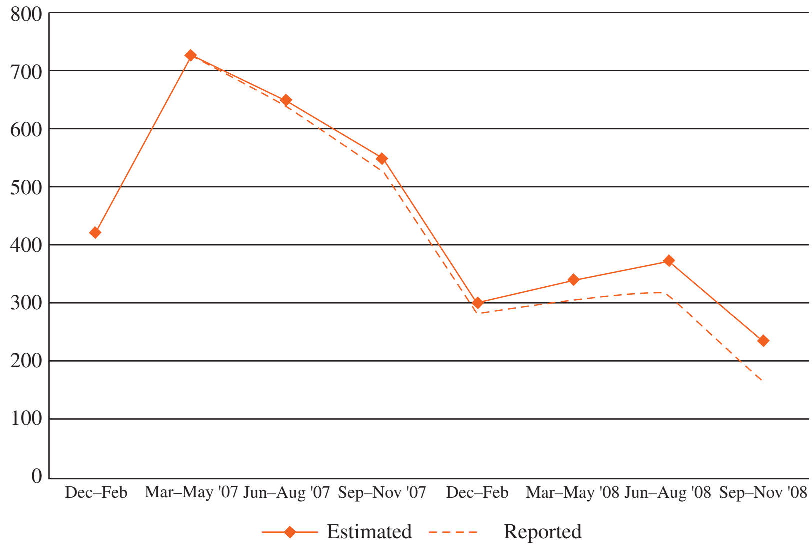


**Figure 1.3: Total loan issuance, US Corporate Loans**

Compiled from DealScan database of loan originals. Reported corresponds to loans reported in DealScan as of December 1, 2008. *Source:* *Ivashina and Scharfstein (2008), Figure 1.*

The financial crisis of 2007–2009 is estimated to have cost the US economy 22 trillion dollars, not only in the form of declines in asset values but also in the form of lost output.[18] Worldwide losses have been estimated to be not less than 1 year's world output, and as much as five times larger (Haldane (2010)). The cost of a malfunctioning financial services industry is clearly very great and, as of this writing, the world economy has yet to recover fully from the financial crisis of 2007–2009.[19]

The "Great Recession" was also the culmination of a large and rapid expansion of financial services along a number of dimensions. As a percentage of US GDP, the financial services sector rose from 4.9% in 1980 to 8.3% in 2006. The value of all United States issued private financial claims (stocks, bonds, etc.) rose from five times US GDP in 1980 to 10 times US GDP in 2007 (Greenwood and Scharfstein, 2013), a phenomenon observed in other well developed countries (see Figure 1.4). The provision of household credit similarly rose from an aggregate value equal to 0.48 GDP in 1980 to 0.99 GDP in 2007.

This enormous expansion in financial activity, in conjunction with its subsequent collapse, has led some to question whether the finance industry has grown too large. Pagano (2012), in particular, makes a strong case in this regard. For a large sample of countries,

[18] Source: General Accounting Office Report #GAO-13-180, "Financial Regulatory Reform: Financial Crisis Losses and Potential Impacts of the Dodd-Frank Act," January 16, 2013.

[19] We give a detailed overview of the 2007 financial crisis in Web Chapter D.

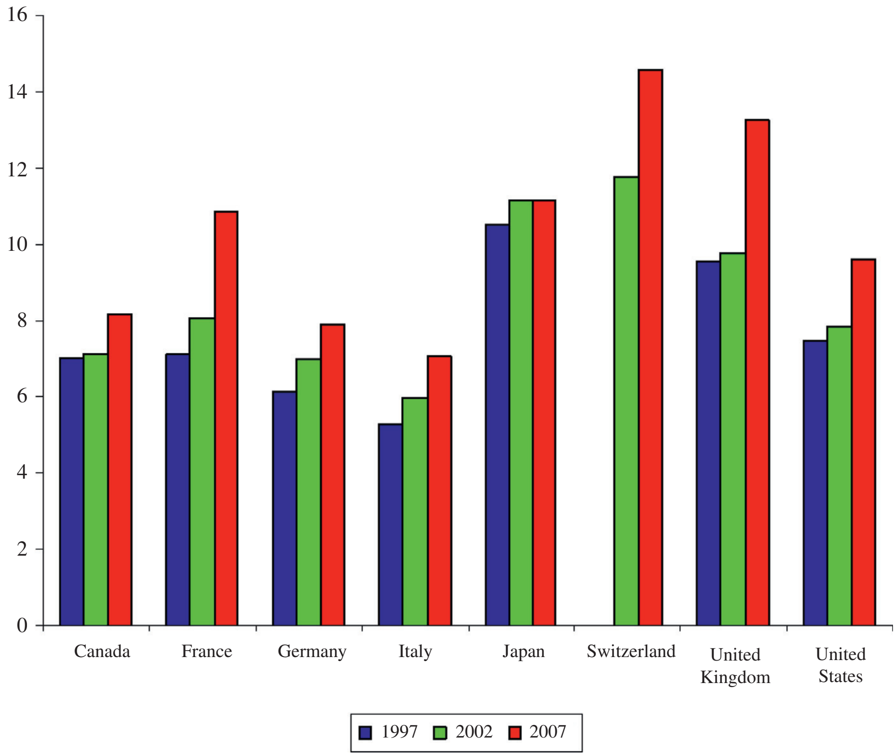


**Figure 1.4: Financial deepening, advanced economies**[21]

Ratio of total financial assets to GDP. *Source: Financial Accounts Statistics, OECD and Eurostat.*

he measures a country's degree of financial development either by the ratio of outstanding private credit to GDP or the ratio of aggregate stock market value to GDP (in both cases averages for the 1980–1995 period). Using these measures, he demonstrates a strong positive association between the growth in value added for industries that are highly dependent on external finance and either measure of financial development in non-OECD countries. For OECD countries, however, the association is small and not statistically different from zero. These results suggest that for countries in the earlier stages of their economic development, an expansion of the financial industry can enhance economic growth, but for countries with well-developed financial markets, this is no longer the case. Pagano (2012) concludes: "Beyond a certain point, financial development does not appear to contribute significantly to economic activity."[20]

[20] Pagano (2012), page 3.

[21] This figure is taken from Milesi-Ferretti and Tille (2010).

He then goes on to study the relationship of bank credit-worthiness (as captured by a special index, which he constructs) and the ratio of private credit issued by deposit banks and other financial institutions to GDP as the measure of financial development. For developing countries where the private credit to GDP ratio is less than 50%, he finds a positive association (correlation) between these quantities; for developed countries with private credit to GDP ratios above 50%, however, the correlation turns negative. It becomes more negative for those countries where private credit as a fraction of GDP exceeds 100%, which is the case for the United States and the United Kingdom. While not formally conclusive, the Pagano (2012) results do suggest that in some countries the financial industry may have grown so large that it now has “a life of its own,” one with self interests that potentially compromise its primary role of matching savers to investors in an efficient way.

## 1.8 Conclusion

To conclude this introductory chapter, we advance a vision of the financial system progressively evolving toward the complete markets paradigm, starting with the most obviously missing markets and slowly, as technological innovation decreases transaction costs and allows the design of more sophisticated contracts, completing the market structure. Have we arrived at a complete market structure? Have we come significantly closer? There are opposing views on this issue. While a more optimistic perspective is proposed by Merton (1990) and Allen and Gale (1994), we choose to close this chapter on two healthily skeptical notes. Tobin (1984, p. 10), for one, provides an unambiguous answer to the above question:

> *New financial markets and instruments have proliferated over the last decade, and it might be thought that the enlarged menu now spans more states of nature and moves us closer to the Arrow—Debreu ideal. Not much closer, I am afraid. The new options and futures contracts do not stretch very far into the future. They serve mainly to allow greater leverage to short-term speculators and arbitrageurs, and to limit losses in one direction or the other. Collectively they contain considerable redundancy. Every financial market absorbs private resources to operate, and government resources to police. The country cannot afford all the markets the enthusiasts may dream up. In deciding whether to approve proposed contracts for trading, the authorities should consider whether they really fill gaps in the menu and enlarge the opportunities for Arrow—Debreu insurance, not just opportunities for speculation and financial arbitrage.*

Shiller (1993, pp. 2—3) is even more specific with respect to missing markets:

> *It is odd that there appear to have been no practical proposals for establishing a set of markets to hedge the biggest risks to standards of living. Individuals and organizations could hedge or insure themselves against risks to their standards of living if an array of*

*risk markets—let us call them macro markets—could be established. These would be large international markets, securities, futures, options, swaps or analogous markets, for claims on major components of incomes (including service flows) shared by many people or organizations. The settlements in these markets could be based on income aggregates, such as national income or components thereof, such as occupational incomes, or prices that value income flows, such as real estate prices, which are prices of claims on real estate service flows.*

## References

Aghion, P., Fally, T., Scarpetta, S., 2007. Credit constraints as a barrier to entry and post-entry growth of firms. Econ. Policy. 22, 731−779.

Allen, F., Gale, D., 1994. Financial Innovation and Risk Sharing. MIT Press, Cambridge, MA.

Arrow, K.J., 1964. The role of securities in the allocation of risk. Rev. Econ. Stud. 31, 91−96.

Barro, R.J., Sala-i-Martin, X., 1995. Economic Growth. McGraw-Hill, New York, NY.

Bernanke, B., Gertler, M., Gilchrist, S., 1996. The financial accelerator and the flight to quality. Rev. Econ. Stat. 78, 1−15.

Bernstein, P.L., 1992. Capital Ideas. The Improbable Origins of Modern Wall Street. The Free Press, New York, NY.

Debreu, G., 1959. Theory of Value: An Axiomatic Analysis of Economic Equilibrium. John Wiley & Sons, New York, NY.

Drèze, J.H., 1971. Market Allocation Under Uncertainty. Eur. Econ. Rev. 2, 133−165.

General Accounting Office Report #GAO-13-180, Financial Regulatory Reform: Financial Crisis Losses and Potential Impacts of the Dodd-Frank Act, January 16, 2013.

Greenwood, R., Scharfstein, D., 2013. The growth of finance. J. Econ. Perspect. 27, 3−28.

Haldane, A., 2010. The $100 Billion Question. Comments given at the Institute of Regulation and Risk, Hong Kong, March 30. <http://bankofengland.co.uk/publications/documents/speeches/2010/speech433.pdf>.

Ivashina, V., Scharfstein, D., 2008. Bank lending during the financial crisis of 2008. Working paper, Harvard Business School. Published under the same title in Journal of Financial Economics, 2010 (97), 319−338.

Jappelli, T., Pagano, M., 1994. Savings, growth, and liquidity constraints. Q. J. Econ. 109, 83−109.

King, R., Levine, R., 1993a. Finance and growth: Schumpeter may be right. Q. J. Econ. 108, 713−737.

King, R., Levine, R., 1993b. Finance, entrepreneurship and growth. J. Monet. Econ. 32, 513−542.

Lee, I.H., Syed, M., Xueyau, L., 2012. Is China overinvesting and does it matter? IMF Working Paper WP/12/277.

Levine, R., 1997. Financial development and economic growth: views and agenda. J. Econ. Lit. 35, 688−726.

Merton, R.C., 1990. The financial system and economic performance. J. Financ. Serv. 4, 263−300.

Milesi-Ferretti, G.-M., Tille, C., 2010. The great retrenchment: international capital flows during the global financial crisis. Working Paper, CEPR, Economic Policy.

Pagano, M., 2012. Finance: economic lifeblood or toxin? Working Paper, CEPR and University of Naples Federico II.

Rajan, R., Zingales, L., 1998. Financial dependence and growth. Am. Econ. Rev. 88, 559−587.

Reinhart, C., Rogoff, K., 2009. This Time Is Different: Eight Centuries of Financial Folly. Princeton University Press, Princeton.

Schumpeter, J., 1934. The Theory of Economic Development, Duncker & Humblot, Leipzig. Trans. R. Opie (1934),Harvard University Press, Cambridge, MA.

Shiller, R.J., 1993. Macro Markets—Creating Institutions for Managing Society's Largest Economic Risks. Clarendon Press, Oxford.

Solow, R.M., 1956. A contribution to the theory of economic growth. Q. J. Econ. 32, 65−94.

Tobin, J., 1984. On the efficiency of the financial system. Lloyds Bank Rev.1−15.

## Complementary Readings

As a complement to this introductory chapter, the reader will be interested in the historical review of financial markets and institutions found in the first chapter of Allen and Gale (1994). Bernstein (1992) provides a lively account of the birth of the major ideas making up modern financial theory, including personal portraits of their authors.

## Appendix: Introduction to General Equilibrium Theory

The goal of this appendix is to provide an introduction to the essentials of general equilibrium theory, thereby permitting a complete understanding of Section 1.6 and facilitating the discussion of subsequent chapters (from Chapter 8 onward). To make this presentation as simple as possible, we will take the case of a hypothetical exchange economy (i.e., one with no production) with two goods and two agents. This permits using a very useful pedagogical tool known as the Edgeworth–Bowley box.

Let us analyze the problem of allocating efficiently a given economywide endowment of 10 units of good 1 and 6 units of good 2 among two agents, A and B. In Figure A1.1, we measure good 2 on the vertical axis and good 1 on the horizontal axis. Consider the choice problem from the origin of the axes for Mr. A, and upside down (i.e., placing the origin in the upper right corner), for Ms. B. An allocation is then represented as a point in a rectangle of size $6 \times 10$. Point E is an allocation at which Mr. A receives 4 units of

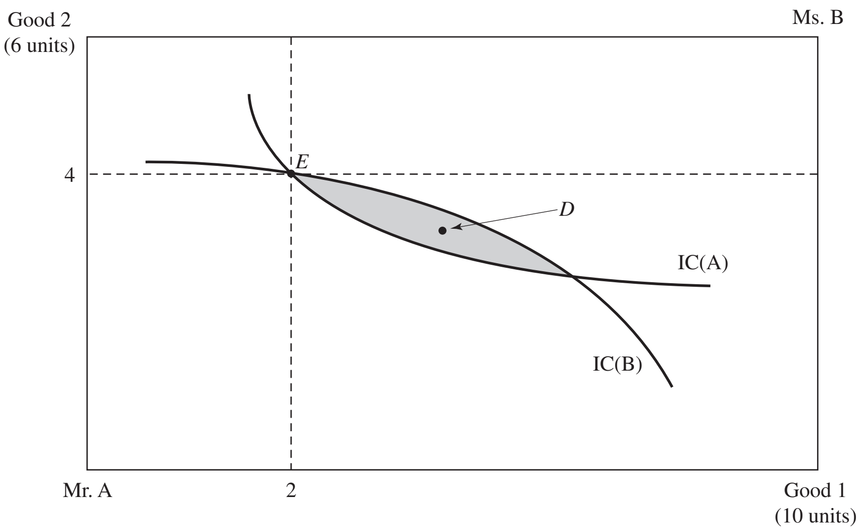


**Figure A1.1**
The Edgeworth–Bowley box: the set of Pareto superior allocations.

good 2 and 2 units of good 1. Ms. B gets the rest, 2 units of good 2 and 8 units of good 1. All other points in the box represent feasible allocations, i.e., alternative ways of allocating the resources available in this economy.

### *Pareto Optimal Allocations*

In order to discuss the notion of Pareto optimal or efficient allocations, we need to introduce agents' preferences. They are fully summarized, in the graphical context of the Edgeworth–Bowley box, by indifference curves (IC) or utility level curves. Starting from the allocation E represented in Figure A1.1, we can thus record all feasible allocations that provide the same utility to Mr. A. The precise shape of such a level curve is person specific, but we can at least be confident that it slopes downward. If we take away some units of good 1, we have to compensate him with some extra units of good 2 if we are to leave his utility level unchanged. It is easy to see as well that the ICs of a consistent person do not cross, a property associated with the notion of transitivity (and with rationality) in Chapter 3. And we have seen in Boxes 1.1 and 1.2 that the preference for smoothness translates into a strictly concave utility function, or, equivalently, convex-to-the-origin level curves as drawn in Figure A1.1. The same properties apply to the IC of Ms. B, of course viewed upside down with the upper right corner as the origin.

With this simple apparatus we are in a position to discuss further the concept of Pareto optimality. Arbitrarily tracing the level curves of Mr. A and Ms. B as they pass through allocation E (but in conformity with the properties derived in the previous paragraph), only two possibilities may arise: they cross each other at E or they are tangent to one another at point E. The first possibility is illustrated in Figure A1.1, the second in Figure A1.2. In the first case, allocation E cannot be a Pareto optimal allocation. As the picture illustrates clearly, by the very

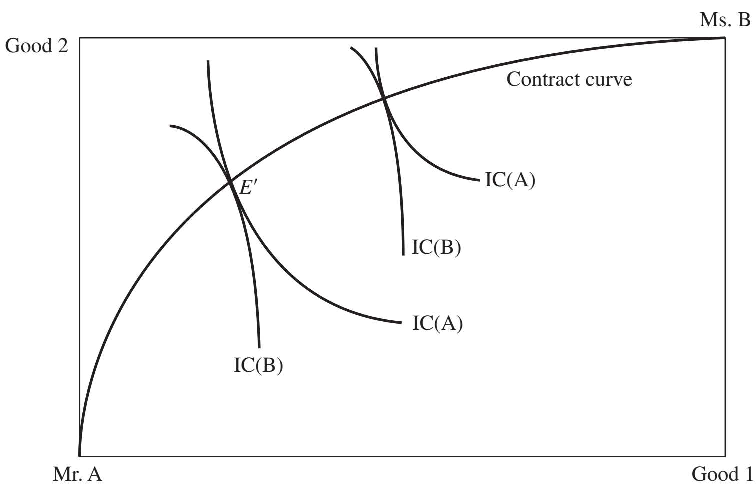


**Figure A1.2**
The Edgeworth–Bowley box: the contract curve.

definition of level curves, if the ICs of our two agents cross at point E, there is a set of allocations (corresponding to the shaded area in Figure A1.1) that both Mr. A and Ms. B simultaneously prefer to E. These allocations are Pareto superior to E, and, in that situation, it would indeed be socially inefficient or wasteful to distribute the available resources as indicated by E. Allocation D, for instance, is feasible and preferred to E by both individuals.

If the ICs are tangent to one another at point E′ as in Figure A1.2, no redistribution of the given resources exists that would be approved by both agents. Inevitably, moving away from E′ decreases the utility level of one of the two agents if it favors the other. In this case, E′ is a Pareto optimal allocation. Figure A1.2 illustrates that it is not generally unique, however. If we connect all the points where the various ICs of our two agents are tangent to each other, we draw the line, labeled the contract curve, representing the infinity of Pareto optimal allocations in this simple economy.

An indifference curve for Mr. A is defined as the set of allocations that provide the same utility to Mr. A as some specific allocation; e.g., allocation E: $\{(c_1^A, c_2^A): U(c_1^A, c_2^A) = U(E)\}$. This definition implies that the slope of the IC can be derived by taking the total differential of $U(c_1^A, c_2^A)$ and equating it to zero (no change in utility along the IC), which gives:

$$\frac{\partial U(c_1^A, c_2^A)}{\partial c_1^A} dc_1^A + \frac{\partial U(c_1^A, c_2^A)}{\partial c_2^A} dc_2^A = 0 \tag{1.3}$$

and thus

$$-\frac{dc_2^A}{dc_1^A} = \frac{\dfrac{\partial U(c_1^A, c_2^A)}{\partial c_1^A}}{\dfrac{\partial U(c_1^A, c_2^A)}{\partial c_2^A}} \equiv \mathrm{MRS}_{1,2}^A \tag{1.4}$$

That is, the negative (or the absolute value) of the slope of the IC is the ratio of the marginal utility of good 1 to the marginal utility of good 2 specific to Mr. A and to the allocation $(c_1^A, c_2^A)$ at which the derivatives are taken. It defines Mr. A's marginal rate of substitution (MRS) between the two goods.

Equation (1.4) permits a formal characterization of a Pareto optimal allocation. Our former discussion has equated Pareto optimality with the tangency of the ICs of Mr. A and Ms. B. Tangency, in turn, means that the slopes of the respective ICs are identical. Allocation E, associated with the consumption vector $(c_1^A, c_2^A)^E$ for Mr. A and $(c_1^B, c_2^B)^E$ for Ms. B, is thus Pareto optimal if, and only if,

$$\mathrm{MRS}_{1,2}^A = \frac{\dfrac{\partial U(c_1^A, c_2^A)^E}{\partial c_1^A}}{\dfrac{\partial U(c_1^A, c_2^A)^E}{\partial c_2^A}} = \frac{\dfrac{\partial U(c_1^B, c_2^B)^E}{\partial c_1^B}}{\dfrac{\partial U(c_1^B, c_2^B)^E}{\partial c_2^B}} = \mathrm{MRS}_{1,2}^B \tag{1.5}$$

Equation (1.5) provides a complete characterization of a Pareto optimal allocation in an exchange economy except in the case of a corner allocation, i.e., an allocation at the frontier of the box where one of the agents receives the entire endowment of one good and the other agent receives none. In that situation, it may well be that the equality could not be satisfied except, hypothetically, by moving to the outside of the box, i.e., to allocations that are not feasible since they require giving a negative amount of one good to one of the two agents.

So far we have not touched on the issue of how the discussed allocations may be determined. This is the viewpoint of Pareto optimality, which analysis is exclusively concerned with deriving efficiency properties of given allocations, regardless of how they were achieved. Let us now turn to the concept of competitive equilibrium.

### *Competitive Equilibrium*

Associated with the notion of competitive equilibrium is the notion of markets and prices. One price vector (one price for each of our two goods), or simply a relative price taking good 1 as the numeraire, and setting $p_1 = 1$, is represented in the Edgeworth–Bowley box by a downward-sloping line. From the viewpoint of either agent, such a line has all the properties of the budget line. It also represents the frontier of their opportunity set. Let us assume that the initial allocation, before any trade, is represented by point I in Figure A1.3. Any line sloping downward from I does represent the set of allocations that Mr. A, endowed with I, can obtain by going to the market and exchanging (competitively, taking prices as given)

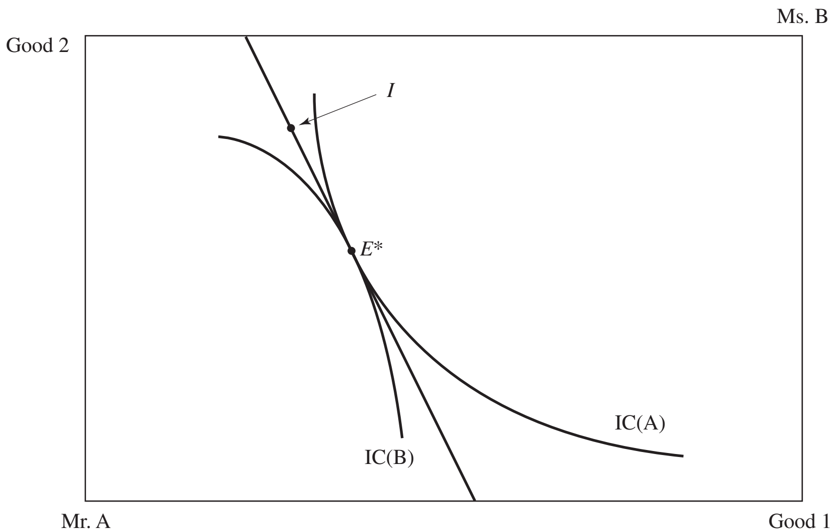


**Figure A1.3**
The Edgeworth–Bowley box: equilibrium achieved at E*.

good 1 for 2 or vice versa. He will maximize his utility subject to this budget constraint by attempting to climb to the highest IC making contact with his budget set. This will lead him to select the allocation corresponding to the tangency point between one of his ICs and the price line. Because the same prices are valid for both agents, an identical procedure, viewed upside down from the upper right-hand corner of the box, will lead Ms. B to a tangency point between one of her ICs and the price line. At this stage, only two possibilities may arise: Mr. A and Ms. B have converged to the same allocation (the two markets, for good 1 and 2, clear—supply and demand for the two goods are equal and we are at a competitive equilibrium); or the two agents' separate optimizing procedures have led them to select two different allocations. Total demand does not equal total supply, and an equilibrium is not achieved. The two situations are described, respectively, in Figures A1.3 and A1.4.

In the disequilibrium case of Figure A1.4, prices will have to adjust until an equilibrium is found. Specifically, with Mr. A at point A and Ms. B at point B, there is an excess demand of good 2 but insufficient demand for good 1. One would expect the price of 2 to increase relative to the price of good 1 with the likely result that both agents will decrease their net demand for 2 and increase their net demand for 1. Graphically, this is depicted by the price curve tilting with point I as the axis and looking less steep (indicating, for instance, that if both agents wanted to buy good 1 only, they could now afford more of it). With regular ICs, the respective points of tangencies will converge until an equilibrium similar to the one described in Figure A1.3 is reached.

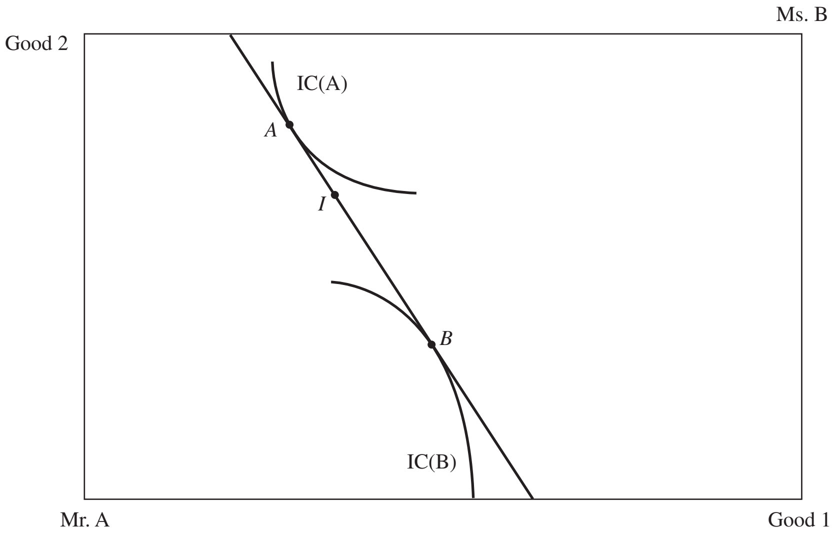


**Figure A1.4**
The Edgeworth–Bowley box: disequilibrium, excess demand for good 2, excess supply for good 1.

We will not say anything here about the conditions guaranteeing that such a process will converge. Let us rather insist on one crucial necessary precondition: that an equilibrium exists. In the text we have mentioned that assumptions H1 to H4 are needed to guarantee the existence of an equilibrium. Of course, H4 does not apply here. H1 states the necessity of the existence of a price for each good, which is akin to specifying the existence of a price line. H2 defines one of the characteristics of a competitive equilibrium: that prices are taken as given by the various agents and the price line describes their perceived opportunity sets. Our discussion here can enlighten the need for H3. Indeed, in order for an equilibrium to have a chance to exist, the geometry of Figure A1.3 makes clear that the shape of the two agents' ICs is relevant. The price line must be able to separate the "better than" areas of the two agents' ICs passing through a same point—the candidate equilibrium allocation. The better than area is simply the area above a given IC. It represents all the allocations providing higher utility than those on the level curve. This separation by a price line is not generally possible if the ICs are not convex, in which case an equilibrium cannot be guaranteed to exist. The problem is illustrated in Figure A1.5.

Once a competitive equilibrium is observed to exist, which logically could be the case even if the conditions that guarantee existence are not met, the Pareto optimality of the resulting allocation is ensured by H1 and H2 only. In substance this is because once the common price line at which markets clear exists, the very fact that agents optimize taking prices as given leads them to a point of tangency between their highest IC and the common price line. At the resulting allocation, both MRS are equal to the same price line and, consequently, are identical. The conditions for Pareto optimality are thus fulfilled.

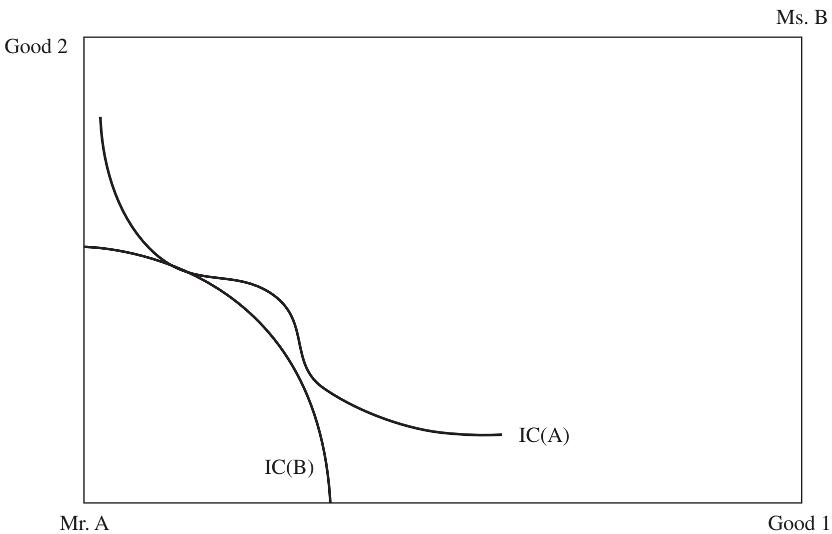


**Figure A1.5**
The Edgeworth–Bowley box: nonconvex indifference curves.