This chapter is incomplete but contains the relevant material you need.

20

Credit 2: Responses to Market Failure

20.1. Introduction

The previous chapter introduced us to some fundamental forces that impede the workings of credit markets. In this addendum, we study a few responses to market failure, driven by informal arrangements or microfinance organizations.

We've already discussed formal and informal credit markets, and we know that a large share of credit in developing countries can be attributed to informal lending. As already noted, a majority of informal lenders do *not* pursue usury as their sole occupation. Many are also wealthy landlords, shopkeepers, or traders. For instance, landlords tend to give credit mostly to their tenants or farm workers, whereas traders favor lending to clients from whom they also purchase grain (see the box on the Philippines). The presence of the production or trade relationship facilitates the credit relationship. This *interlinkage* — business conducted across the same trading partners, with the terms of transaction in one market depending on those in the other — is a far cry from the impersonal and independent functioning of markets that characterizes most textbook economic theory.

A second, very important response to the imperfection of credit markets is the rise of microfinance organizations. The pioneering organization came out of Bangladesh — the famous Grameen Bank started by Mohammed Younus. Microfinance refers to the financing of individuals or small businesses at tiny scales that a large bank would not even tough. They advance loans, often for working capital, but sometimes also to finance a startup business. Among the innovative features of microfinance organizations — again pioneered by Grameen — is the use of frequent repayment schedules involving very small sums of money, a practice that is meant to inculcate disciple and a sense of accomplishment among the borrowers as they see their debt whittled down slowly but steadily. Microfinance organizations have also experimented with *group liability*, a more controversial practice in which groups of individuals is held jointly liable for their debts.

The final response we study is mutual credit and insurance based on reciprocity across similarly located individuals in a community, backed by the threat of group sanctions in the event of default. The ability to implement such sanctions on the

appropriate occasion relies, of course, on group information. about its members. Consider crop insurance. If such insurance is provided by a large company without detailed knowledge of the insurees, it is hard to condition the terms of the contract on appropriate information. There are, of course, some observable shocks such as weather that the company can condition its payments upon, but very often there are farm-specific, idiosyncratic shocks as well that the company cannot observe. Additionally, production *inputs* cannot be verified. For instance, crop insurance might encourage undersupply of effort on the farm, under the expectation that a bad crop will be insured.

It is not surprising, then, that insurance schemes often arise within groups of people who co-insure one another under better informational conditions. As a farmer and a village resident, it may be difficult for my next-door neighbor to convince me that his crop has failed when I know about his activities on a daily basis. I may also be able to monitor his effort better, and he mine. However, such schemes, better grounded in information though they may be, can obviously not provide insurance against a correlated event (think of weather conditions such as drought or flooding).[298] Group sanctions notwithstanding, they also continue to suffer from a problem analogous to strategic default. An individual in good economic shape this year may be that much less willing to help out someone who is not. He knows that in displaying this unwillingness, he may be locked out of future access to insurance (and he may face the disapproval of the community as well), but this may or may not be enough to deter him.

20.2. Interlinked Contracts

A common feature of many loan transactions in developing countries is that credit is linked with dealings in some other market, such as the market for labor, land, or crop output. For instance, it is commonly observed that landlords are often the principal source of credit for their tenants, using their labor or even their rights to tenancy as some form of collateral. On the other hand, traders are the principal source of funds to owner–cultivators, especially those who lack access to the formal sector. Traders usually combine such credit dealings with purchase of their borrower's crop.

Interlinked contracts are formed in a variety of other ways. While loans are outstanding, the lender may have use rights to the land or other assets of the borrower, as already discussed. To the extent that the lender can directly benefit from the sorts of assets owned by the borrower, this makes credit transactions easier to enforce. However, for that direct benefit to be present, it is often necessary that the borrower and lender be engaged in similar or complementary occupations.

In one sense, interlinkage is just a marriage of convenience. If a lender also has a principal occupation that ties in well with the occupation of the borrower, it may simply be convenient to carry out credit and other dealings under one umbrella contract, explicit or implicit. Thus a trader who transports rice may also advance credit to a rice-growing farmer, as well as trade in the rice produced by the farmer. In this marriage, there isn't really any synergy between the two activities, except that they both *happen* to be carried out by the same pair of economic agents. Neither half of the deal rests tightly on the other, and in this sense the transactions are not linked.

[298] Large companies can handle correlated risk, but must be diversified over a larger area. Such diversification does not, however, allow for microknowledge of conditions within an individual village.

Whatever the ultimate explanation, it does appear that in the event of coincident occupations, the interlinking moneylender has an edge over other moneylenders in credit dealings. In many parts of the developing world, the "pure moneylender" is a dying breed. Individuals engaged in moneylending most likely have a principal occupation that is not moneylending.[299]

In their study of interlinked contracts in the Philippines, Floro and Yotopoulos (1991) distinguished between various forms of interlinkage:

> Five types of interlinkage are distinguished, depending on whether the loan is tied to (1) the provision of intermediation services in relending and/or procuring output; (2) the sale of output to the lender; (3) the purchase of inputs or lease of farm equipment from the lender; (4) transfer of rights over the usufruct of the land to the lender; and (5) the provision of labor services to the lender. [Our study] indicates that the first three types are prevalent among trader–lenders, while the two last among farmer–lenders.

20.2.1. Trade-Credit Interlinkage. We illustrate an interlinked relationship using a trade-credit contract. Danilo grows rice and sells it through a competitive trader, María, thereby earning the market price p net of Maria's trading costs that we ignore here.[300] Danilo needs working capital to grow his rice, just as in Section 19.6: a loan of l generates rice output $Y = F(L)$.

Now suppose that María also has access to loanable funds at some opportunity rate of interest i, and can lend to Danilo. To fix our ideas, suppose first that María simply acts as a pure moneylender; that is, she chooses a rate of interest r at which she lends to Danilo, and Danilo chooses how much to borrow. That is, we're temporarily operating under the "marriage of convenience" view that trading and borrowing are two separate transactions that just happen to be carried out by the same pair. Then, given r, Danilo will choose L to maximize his net payoff, which is

$$\text{Borrower Payoff} = F(L) - (1+r)L.$$

The obvious "marginal-benefit-equals-marginal-cost" calculus shows that $pF'(\hat{L}) = 1+r$ at Danilo's optimal choice, which we've denoted by $\hat{L}$.[301] Figure 20.1 illustrates this outcome. It also depicts María's net payoff:

$$\text{Lender Payoff} = (r-i)\hat{L}.$$

The *total* payoff generated from this interaction, or the social surplus, is therefore

$$\begin{aligned}\text{Borrower Payoff} + \text{Lender Payoff} &= F(\hat{L}) - (1+r)\hat{L} + (r-i)\hat{L} \\ &= F(\hat{L}) - (1+i)\hat{L}.\end{aligned}$$

Note again that the social surplus treats all loan repayments as an inter-person transfer and so these do not directly enter into the social surplus, *except insofar as the loan terms affect the loan chosen*. Figure 20.1 displays the rate of interest $r > i$ chosen by María,

[299] Interlinkages need not be linked to credit transactions alone, although they have been studied most often in this context. For an example of land-labor interlinkage, see Sadoulet's (1992) study on *inquilinaje* in Latin America.

[300] This example is based on Gangopadhyay and Sengupta (1987).

[301] This assumes, of course, that the *borrower* can freely chosen the loan size while the lender dictates the terms. We return to this point below.

and Danilo's reaction $\hat{L}$. At that loan size, Danilo's payoff is depicted by the vertical gap between output value and the repayment of loan (with interest), while María's payoff is given by the gap between the latter and *her* cost of loan provision, which is $(1+i)L$. Adding the two yields the surplus generated by their credit arrangement, which as we've already seen is $F(\hat{L}) - (1+i)\hat{L}$. But Figure 20.1 also illustrates that Danilo's and María's joint surplus would have been maximized at L^*, which satisfies the condition $pF'(L^*) = 1+i$. The very fact that María attempts to set $r > i$ to make a profit *also* implies that the social surplus is distorted below the maximum surplus available.

If there were only some way in which María could *force* Danilo to take a loan of L^* and somehow pry out some of that surplus for herself! You can easily see from the Figure that this would make María even better off, while retaining the same payoff for Danilo. As we've just argued, there's room for it, because the sum of payoffs is certainly larger at L^* than it is at $\hat{L}$. Well, leaving out the considerations of default that we've emphasized in Chapter 19, she most certainly could push the higher loan L^* as an all-or-nothing contract. The problem is that that would be taking our simplified model all too seriously. María doesn't know Danilo's local production conditions as well as Danilo does, which make it very hard for her to calculate that value of L^* and impose it on Danilo.[302]

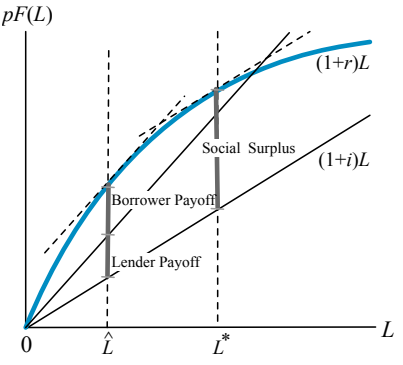

Figure 20.1. Pure Loan Contract.

As it turns out, though, there is an interlinked contract which allows Maria to have her cake and eat it too, one in which Danilo retains full autonomy in choosing the size of the loan. The idea is to depress both Danilo's marginal cost of production and the price he receives for his output, but to do this in equal proportion so that the *ratio* of price to marginal cost is left unchanged relative to $p/(1+i)$. Such a crafty move is akin to a profit tax on Danilo. Danilo will *voluntarily* choose the surplus-maximizing loan size L^*, and María will make the higher payoff she seeks. To make this slightly more precise, define a special "buying price" q^* and an interest rate r^* by two conditions. First, this pair leaves the ratio of price to marginal cost unaffected, as just discussed:

$$q^*/(1+r^*) = p/(1+i). \tag{20.1}$$

Second, by scaling q^* and $1+r^*$ up or down in perfect tandem while maintaining the ratio of the two unchanged as in equation (20.1), the pair (q^*, r^*) yields the same payoff to Danilo as he received under the original pure credit contract:

$$\max_L q^* F(L) - (1+r^*)L = \text{Danilo's payoff under pure credit contract.} \tag{20.2}$$

It is easy to see that there is a unique pair satisfying these two properties. It *must* involve $q^* < p$ and $r^* < i$. This contract must maximize María's profits. The reason is simple and fully illustrated by Figure 20.2. Given condition (20.1), Danilo will choose a loan size exactly equal to L^*, which is the surplus-maximizing amount. By condition (20.2), Danilo is given just enough so that he is indifferent between this and a pure

[302] Ray and Sengupta (1989) contains a detailed discussion of such issues.

credit market. María skims off the remaining surplus, which must generate strictly more profit for her relative to that under the pure-credit contract.

Take note of a remarkable feature of the optimal contract. María actually advances the loan at a rate r^* *below* her opportunity cost of funds, given by the interest rate i. She makes a loss on the credit component of the contract. All her money is made by depressing her buying price q below the market price p for rice. The credit and trading components of the contract cannot be understood in isolation; they are truly interlinked. An economist studying the credit relationship between María and Danilo might be forgiven for thinking that they were the best of buddies — after all, María is apparently subsidizing Danilo's access to credit. But under the hood of the interlinked contract, there is more going on: trade and credit cannot be separated.

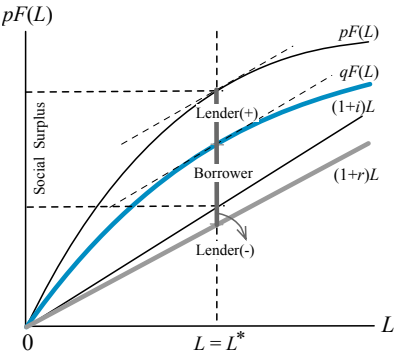

Figure 20.2. Interlinked Contract.

20.2.2. Labor-Credit Interlinkage. Here's a similar argument that establishes the possibility of interlinkage in the credit and *labor* markets.[303] A rural laborer, Anka, must feed herself and her family through both the slack and the peak seasons of an agricultural year. To make the exposition as simple as possible, imagine that there is no employment available in the slack season, whereas in the peak season, harvesting jobs are available that pay a wage of w. To finance her consumption in the slack season, Anka must borrow.

Suppose, now, that there is a large farmer, Birju, who hires harvesting labor during the peak season and has access to funds at an opportunity cost of i per unit, just as María the trader had in the previous section. Birju is in a position to lend money to Anka. He could do that by offering an interest rate r on loans, with Anka then choosing how much to borrow. If we denote Anka's slack consumption by c_s and her peak consumption by c_p, then Anka's budget constraint is given by

$$w = c_s(1 + r) + c_p, \tag{20.3}$$

and she would choose the two consumptions within this constraint to make herself as well off as possible. Figure 20.3 illustrates. The blue curves are Anka's indifference curves defined on slack and peak consumption. The intercept w shows her wage, and the line that connects w to a show the various combinations of c_s and c_p that are available to her at the interest rate r. She choose a loan of $\hat{c}_s$ to maximize her utility. The corresponding profit to Birju is shown by the thick vertical segment that chalks out the difference between the repayment from Anka, given by $\hat{c}_s(1 + r)$, and his opportunity cost, equal to $\hat{c}_s(1 + i)$. The latter is depicted for various values of c_S by the line $[w, d]$, which slopes down at a gradient with absolute value equal to $1 + i$.

This is a nice healthy profit as far as Birju is concerned but it can be made healthier. An alternative arrangement can be found that keeps Anka on her achieved indifference curve, and gives Birju still higher profit. It involves the charging of an interest rate of

[303] The analysis in this section follows Bardhan (1984).

just i to Anka, but asking her to work at the reduced wage w^* instead of w. Anka's constraint set is now given by the line segment $[w^*, b]$, or all combinations of slack and peak consumption such that

$$w^* = c_s(1+i) + c_p, \tag{20.4}$$

The wage w^* is chosen so that Anka achieves the same utility as before by choosing to consume c_S^* in the slack season. Birju makes no money off the loan, *but gets to hire Anka at a reduced wage*, and makes a larger overall profit as a consequence.

The intuition behind this result is very close to the one given in the trader-lender example. The sum of payoffs to borrower and lender are maximized at a loan equal to c_S^*, which would be chosen were Anka herself able to borrow from the formal credit market. If Anka can be kept on the same indifference curve as under pure moneylending, then Birju gets to claim all the rest of the surplus. The problem is that with a pure credit contract, the higher interest rate r distorts the loan Anka is willing to take, leading to a drop in the sum of payoffs. By asking for repayment in reduced wages, the joint contract neatly avoids this distortion.

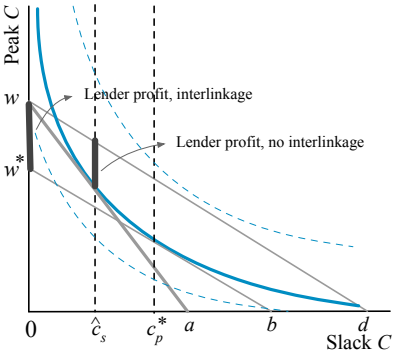

Figure 20.3. Labor-credit interlinkage.

The dominant contract, then, is indeed an interlinked transaction. No extra interest is charged on the loan; all payments are made in "labor units." As in the trader-lender story, these are not just two independent contracts. If Birju offered Anka a wage below w, and no loan, Anka would not take it. Likewise, if Birju were restricted to offer pure moneylending contracts, he would push the interest rate above i. It is really a combination of a labor and credit contract, the two components of which depend on each other and cannot be offered separately.

20.2.3. Other Approaches to Interlinkage. So far, we have relied on a particular view of interlinkage: that rorrower-lender interaction can distort relative prices, thereby reducing their joint payoffs. If borrower and lender have no overlap on other markets, there is little that can be done about the distortion. From the lender's perspective, a positive share of a smaller cake is better than a minuscule share of a larger cake. But if there is simultaneous interaction across distinct markets, the distortion can be reduced through interlinkage, as we've shown in the scenarios just discussed.

Narayana Kurup's (1976) study of Kerala, India, is instructive in this respect:

> "[T]he loan transactions observed in this area seem to fall into certain clear-cut categories. They are loans based on mortgage of land or usufructs, loans given by coconut traders to cultivators on the understanding that the latter would sell the coconuts to their creditors, and loans advanced to agricultural laborers who undertake to sell their harvest wages (i.e., paddy) to their creditors. The practice of taking loans by mortgaging the ration cards is also found to be quite common in this locality … There is no explicit interest involved; the vast majority

> of borrowers do not believe that they are paying any interest. But the usufructs that the creditors enjoy constitute the interest…
>
> The foregoing discussion has shown clearly that there exists, in this area, a variety of loan transactions which, on surface, are interest-free but where heavy interest is in fact hidden under the rug. Since the implied rates of interest appear in many disguised forms, even the borrowers do not recognize the fact. No wonder, the presence of such implicit interest charges entirely escapes the net of conventional surveys."

Kurup's study is entirely consistent with the idea of reducing distortions, and confirms that *explicit* interest rates could be relatively low under an interlinked contract; see also Floro and Yotopoulos (1991). Kurup also notes that that absence of interest is deceptive: with transactions, interest may be hidden in other features of the overall deal (such as the price at which a trader buys output from the farmer or the implicit wage at which a laborer is required to work off an ostensibly interest-free loan).[304]

The goal of reducing distortions may be one reason for low interest rates. But low rates are also particularly convenient for societies in which the explicit charging of interest is forbidden or shunned. In Islamic societies, *riba*, or the charging of "excessive interest," is generally regarded as immoral and so is banned under the Shaariat law, though the quotation marks are meant to suggest some ambiguity about the term. In such situations, it is prudent to ask for interest in secondary forms and advance the loan interest-free. Interlinked contracts provide a way out. A large landowner may ask for the rights to part of the borrower's output as long as the loan is outstanding, even though the loan is recorded as interest-free. A trader may make no-interest cash advances to his suppliers, provided that the supplier agrees to sell him the crop at a discount. Even the right to use a ration card, which permits access to the public distribution system of food grain at subsidized prices, constitutes hidden charging of interest while the loan is outstanding. All these contracts may be acceptable under a law or norm that bans *explicit* usury. As already noted, that absence of interest is deceptive: with transactions, interest may be hidden in other features of the overall deal (such as the price at which a trader buys output from the farmer or the implicit wage at which a laborer is required to work off an ostensibly interest-free loan).

This is particularly convenient for societies in which the explicit charging of interest is forbidden or shunned. In Islamic societies, *riba*, or the charging of "excessive interest," is generally regarded as immoral and so is banned under the Shaariat law, though the quotation marks are meant to suggest some ambiguity about the term. In such situations, it is prudent to ask for interest in secondary forms and advance the loan interest-free. Interlinked contracts provide a way out. A large landowner may ask for the rights to part of the borrower's output as long as the loan is outstanding, even though the loan is recorded as interest-free. A trader may make no-interest cash advances to his suppliers, provided that the supplier agrees to sell him the crop at

[304] It isn't that informal interest rates are invariably low. Such rates vary by geographical location, the source of funds, and the characteristics of the borrower. Irfan Aleem's survey (1992) of the Chambar region of Pakistan showed that the average annual interest rate was as high as 78.7%, though it involved substantial dispersion. Siamwalla et al. (1993) reported that in most parts of Thailand, the informal-sector interest rate varied between 5 and 7% *per month*, which is dramatically higher than the 12% *per annum* charged by formal-sector banks. So it is not that interest rates are always low, but that they often *appear* to be that way in interlinked transactions.

a discount. Even the right to use a ration card, which permits access to the public distribution system of food grain at subsidized prices, constitutes hidden charging of interest while the loan is outstanding. All these contracts may be acceptable under a law or norm that bans *explicit* usury.

Additionally, an interlinked arrangement allows the lender to dispense with, or at least consolidate, the costs of tracking her borrower's activities. A rice trader who makes funds available to a farmer may demand repayment in terms of the output because such repayment is easier to enforce under the normal routines of the trader-lender. At harvest time, the trader might arrive at the fields of his suppliers to pick up the crop for transportation. If this is something the trader has to do *anyway*, a useful by-product is that he gets to place first claim on the crop. Such claims can be extremely powerful, effectively pushing other creditors to a secondary position. Likewise, a laborer or tenant farmer who works on the estate of a large landowner under normal circumstances presents a relatively economical credit prospect. In the case of a default on the monetary terms of the loan, the loan can be worked off (and the implicit wages deducted as payment). These are all ways to reduce the chances of involuntary default without having to incur the *extra* costs of monitoring or tracking.

Interlinked relationships are also useful in the prevention of strategic default. Recall the model of default that we considered in Chapter 19.6. We observed that the moneylender could not drive the borrower down to his participation constraint: a surplus over the next best option had to be provided. The borrower trades off the loss in this surplus at future dates with the short-term gain from default. In an interlinked relationship, the lender has other means at his disposal. With more than one point of interactions across the parties, there are multiple sticks that can be used to enforce to reinforce the carrot of any one relationship.

For instance, suppose that a landlord lends money to a tenant who also rents out part of the landlord's land. The repayment of that loan can then be doubly enforced by an "interlinked threat": if the loan is not repaid, then the *tenancy* will be removed. The provision of a valuable tenancy thus serves a twin role. It assures the supply of appropriate tenant effort on the rented land via the threat of eviction,[305] while at the same time it doubles as an incentive to repay loans (see Mansuri 2007). In this sense, the landlord is at a distinct advantage in advancing credit to his tenant, because he has at his disposal a preexisting instrument of repayment. In contrast, a pure moneylender who lends to the same tenant must offer additional incentives for repayment *through the credit contract itself*.

20.3. Alternative Credit Policies

There has been a growing realization that the needs of rural credit cannot be adequately served with the use of large financial institutions such as commercial banks. As already discussed, the microinformation that is required for these operations precludes efficient market coverage on the part of these large organizations.

Two kinds of policies can arise in response to this observation. One is to recognize explicitly that informal lenders are much better placed to grant and recover loans from small borrowers than formal institutions are. The idea, then, is not to try to replace this

[305] See, for instance, Dutta, Ray, and Sengupta (1989), Banerjee, Gertler and Ghatak (2002) and Banerjee and Ghatak (2004).

form of lending, but to *encourage* it by expanding formal credit to economic agents who are likely to use these funds in informal markets. The second approach is to actually design credit organizations at the microlevel that will take advantage of local information in innovative ways. We discuss these two avenues.

20.3.1. Vertical Formal-Informal Links. *Expansion of formal credit to informal lenders* The lending of formal funds to informal markets is not a new phenomenon by any means. Large landowners or traders are in a much better position to put up collateral: from the point of view of the banks, they are good credit risks. They then use the funds to cash in on their informational advantage in informal markets. We have already seen evidence of this in our Philippine and Pakistan examples. The expansion of formal sector credit to these agents generates competition among them, and this hopefully improves the borrowing terms faced by individuals who fall outside the ambit of the formal credit system. Similarly, loans might be made to cooperative groups who are better placed to lend the funds because of social or religious ties among the members.

We can think of this policy as forging "vertical" links between the informal and the formal credit systems, rather than the more traditional policy of "horizontally" displacing one system by another.[306]

Vertical Formal–Informal Linkages: The Philippines

The Philippine government has made active attempts to incorporate the informal sector into the overall strategy of agricultural development.[a] In a sense, efforts to institutionalize the informal sector in the Philippines took place as early as the 1950s with the formation of family-run rural unit banks. Small in their capital base and in their mode of operation, the rural banks were then perceived as the stalwarts of agricultural credit delivery. Matters were not so simple, however. Von Pischke [1991] pointed out, however, that these rural bankers were "prominent local citizens who were often involved in moneylending. By becoming rural bankers they could obtain some capital from government to supplement their own. They could expand their moneylending operations by soliciting deposits." As it turned out, rural banks did not reach the majority of the small farmers.

In recent years, the focus on informal lenders has been strengthened. The government has employed a number of approaches to reach small rural borrowers in an indirect fashion. The rediscounting policies of the Central Bank have played a key role, along with fiscal policies that allocate budgetary resources and external borrowings for special credit schemes. These schemes involve the strengthening of linkages between the formal financial institutions and informal intermediaries such as traders.

However, traders are not the only type of informal intermediaries involved. In 1987, the Land Bank of the Philippines launched an aggressive wholesale lending program at below-market rates of interest to credit cooperatives, in an attempt to reach and serve small farmers (with holdings up to 5 hectares). In 1991 and 1992, respectively, 500,000 and 700,000 families received loans amounting to totals of $270 and 320 million. These amounts were disbursed to cooperatives (3,000 and 6,100 in 1991 and 1992, respectively).

Programs sponsored by the government explicitly use informal lenders as conduits of formal sector funds. In 1984, a government-sponsored program aimed at providing credit at concessionary rates to the agricultural sector by using informal lenders as

[306] For analyses of the "horizontal" scenario, see, for example, Bell [1993] and Kochar [1991].

intermediaries was launched. This scheme, called the National Agricultural Productivity Program (NAPP), comprised twelve specific commodity programs geared toward food self-sufficiency. Informal lender–beneficiaries included traders, millers, and input dealers. Commercial banks and rural banks participated in the program by acting as financial intermediaries between the government and the informal lenders. In addition, the government (through the National Food Authority) was also a direct lender.

Two low-cost special financing programs that are largely targeted to agricultural input suppliers are the Planters' Product Credit Scheme and the End Users/Input Suppliers Assistance Scheme. Under these arrangements, end users and input suppliers receive credit at concessionary rates of interest under the condition that they extend credits (in the form of fertilizer and pesticide credits or cash production loans) to farmers.[b]

Traders and rice millers also were utilized as credit channels under the Intensified Rice Production Program. The term "trader–millers" refers to rice trader–millers accredited by the Quedan Guarantee Fund Board who possess primary facilities such as threshers, driers, and mills. Farmers contract to supply the trader–millers with a specified volume of their produce at a buying price not lower than the government support price. Likewise, the trader–millers are required to enter into a "payment-in-kind" agreement with the National Food Authority where they deliver the milled rice equivalent of the due loan payment. This "promotion of linkages combines the strength of both sectors to supplement the resources of the informal sector." (Ghate [1992, p. 867]).

"Horizontal" competition from the informal sector therefore does not seem to be a threat to the formal institutions nor vice versa. A senior official of one of the largest commercial banks in the Philippines [in an interview with Maria Floro (May 1990)] typifies the viewpoint of the dominant group in the formal financial sector: the commercial bankers. According to this view, informal lenders are not bank competitors. Rather, they complement the lending activities of the banking sector. According to the official,

Vertical Formal–Informal Linkages: The Philippines

The informal lenders operate in a different segment of the market. The bank's market lies largely in corporate and commercial accounts which are generally fully collaterized. In fact, some of the informal lenders are, in effect, conduits of bank funds. This is true with our big grain miller and trader clients who provide advances to the paddy farmers in the regular course of their business.

[a] I draw here on Floro and Ray [1997].

[b] Esguerra [1987] and Floro and Yotopoulos [1991] discuss this in more detail.

By allowing the traders, input suppliers, and millers to participate in the lending program, the NAPP merely formalized what was already commonplace in the informal credit market: the use of contracts that interlink credit provision with specific transactions in other markets.

Possible effects Do these schemes intensify competition among informal lenders? I am not aware of serious empirical work that has addressed this important question. Theoretical analysis suggests, however, that the answer is likely to be mixed. Here are three reasons why.

Costs of monitoring. It is true that intensified competition per se tends to depress the interest rates in informal markets, but there are indirect effects to be considered. Among the most important of these are the fixed costs of lending: the costs of administering a loan, clearing credit histories, and then tracking the loan to make sure it is repaid.

Why are these costs there in the first place? As we have seen, they must be incurred to make sure that the borrower is a good credit risk and to ensure that a loan, once taken, is repaid.

Now, if formal credit is expanded, it is likely to increase the number of active moneylenders, which increases the alternatives available to borrowers. All other things being equal, this raises the probability of a default on an existing loan (because alternative sources are more abundant). To avoid this, a lender will have to expend more resources on tracking a loan and will also have to spend more time and money to clear a borrower's credit history. If an Aleem-type world is a good depiction of reality—one in which lenders largely break even because of monopolistic competition—the increased administrative costs will *raise* the equilibrium rate of interest on a loan. Thus intensified competition can be counterproductive from the point of view of the borrowers. This is the theme explored by Hoff and Stiglitz [1998].

Collusion. It may be that lenders are engaged in collusive transactions; that is, they implicitly or explicitly agree to not invade each other's territory, thereby giving them local monopolies in their spheres of influence. For instance, two traders serving a village or a group of villages might agree to partition potential borrowers in the village into two groups, where the partition depends, perhaps, on the type of crop grown or on other characteristics that make each group relatively attractive to one of the traders. Ex post, each trader will have the incentive to invade the territory of the other (perhaps by offering a slightly better credit contract), but these invasions are kept in check by threats of a "credit war"—a temporary or permanent situation in which the traders compete ferociously with each other as punishment for the initial deviation. This is the sort of repeated game argument that reappears several times in the book [see Appendix 1 (at the end of the book) on game theory for an introduction to the subject]. Collusion is a form of farsighted interaction that is kept in place by the threat of reverting to myopic competition if the pact is broken.[307]

The probability of a strategically sustained cooperative outcome therefore depends on (1) the additional profits initially gained from undercutting one's rival, (2) the subsequent loss in terms of a reduction in profits that occurs as a result of a counterinvasion by the rival trader, and (3) the lag or detection time that elapses between the invasion and the reprisal. The probability for sustained cooperation is higher if the initial gain for invasion is small, or the subsequent loss from counter-invasion is large, or if the time lag between invasion and reprisal is long.

Now, an expansion of formal sector credit has two opposing effects: on the one hand, it tends to expand competition because each trader now has an incentive to undercut his rival. On the other hand, it tends to reinforce collusive practices by increasing the severity of credible punishments when deviations occur. The net effect depends on which of these two forces dominates.

There are actually good reasons to argue that the possibility of strategic cooperation may indeed *improve* with the credit expansion. For instance, a larger invasion of rival territory is easier to detect. Thus invasion gains will not go up as quickly as potential

[307] Again, there is no hard evidence whether lenders actually interact in this fashion (and no evidence against it, either), but there is certainly evidence of collusion among traders in their marketing behavior: see, for example, Umali [1990].

punishments, and this will promote strategic cooperation. Essentially, the *ratio* of invasion profits to punishment losses declines.

Other variations of the model yield similar results. For instance, consider a situation where trader 1 specializes in processing a particular crop A, whereas trader 2 specializes in crop B. Imagine a (proposed) collusive outcome whereby trader 1 (respectively 2) lends money to small borrowers specializing in the production of crop A (respectively crop B). Now trader 1, say, offers a loan to a borrower specializing in crop B on the understanding that the farmer will divert some of his land to the production of crop A, which is then sold to the trader. Suppose that the production function for the farmer exhibits diminishing returns in the production of each crop (although overall it may be more suited to the production of crop B). Then the outcome will be a large gain for trader 1 if the amount of the invasion is small, but with diminishing marginal gains as the size of the intervention becomes larger.

Although the deviation gains are small, the *loss* that can be imposed on the deviating trader following a reprisal can still be substantially large. As his clients shift the pattern of crop production, this leads to a fairly small but positive marginal gain for the rival, but it still imposes large marginal losses on the incumbent trader. In this scenario, the expansion of credit lines to traders also leads to a decrease in the ratio of invasion profits to reprisal losses, leading to an increase in the chances of collusion.

The foregoing examples emphasize the importance of understanding the nature of interaction between lenders. The impact of any (credit) policy initiative on the welfare of borrowers crucially depends on the resulting effects on competition (or collusion). This line of reasoning is pursued by Floro and Ray [1997].

Differential information. Bose [1997] explores a third route. Lenders may have differential information regarding borrowers. For instance, lender A may be able to easily distinguish between good and bad credit risks, whereas lender B cannot. Then lender A essentially contaminates the pool for lender B, because he siphons off the good borrowers. Of course, if A faces a credit ceiling, he cannot siphon off *all* the good borrowers. Thus lender B faces a mix of borrowers. The mix might still be profitable enough for lender B to operate. Now an expansion of formal-sector credit will allow lender A to siphon off even more good borrowers. This might contaminate the available pool to lender B to the extent that he no longer finds it profitable to operate. The increase in A's activities may be more than offset by the shutdown in B's lending.

Here is a numerical example for the sake of concreteness. Suppose that there are 100 borrowers in the economy and each borrower seeks to borrow \$1,000. The opportunity cost of funds for each lender is 10% (say this is the formal interest rate). Each lender charges 50% on informal loans.[308] Suppose that each lender has a credit limit on funds and denote this limit by x, the total number of borrowers that each lender can lend to at the rate of \$1,000 each.[309] A formal-sector expansion of credit is captured by an increase in the value of x.

[308] I am going to assume for simplicity that the rate of interest on informal loans is fixed or at least has a ceiling. This might occur, for instance, if no borrower will want to borrow (or will be able or willing to repay) if the interest rate crosses the ceiling. A variation with endogenous interest rates is easy to construct.

[309] So the value of the credit limit is $1{,}000x$.

Suppose that 80 of the 100 borrowers are "good" and the remaining 20 are "bad." Good borrowers will repay their loans: that is, they will return \$1,500. Bad borrowers default; they return nothing. Lender *A* can figure out who the good borrowers are, but lender *B* cannot.

If the credit limit x is less than 80, lender *A* will use all his resources to lend to good borrowers. Assume that this is the case. Then lender *B* faces a population of $100 - x$, among whom $80 - x$ are good (because lender *A* will have made off with x good borrowers). So the expected value of a loan is given by $(80 - x)/(100 - x)1{,}500$, which is the probability of repayment times the full amount, plus the probability of default times zero. Lender *B* will therefore be willing to lend as long as

$$\frac{80 - x}{100 - x}1{,}500 \geqslant 1{,}100$$

where the right-hand side of this inequality is the opportunity cost of the lender. This inequality will be satisfied as long as x is no greater than 25; otherwise it will fail. This means that if x is expanded by means of a formal-sector policy to increase credit lines, both lenders *A* and *B* will expand their lending activities, but as x crosses 25, lender *B* will shut down altogether. Thus around this mark, formal-sector expansion will lead to a decline in informal-sector lending.

A bottom line All this is not to say that an expansion of credit from the formal sector is bound to fail, but the simplistic expectation that such expansion will invariably lower interest rates for small borrowers may not be automatically upheld. Empirical valuation of such schemes should keep these potential pitfalls in mind.

20.3.2. Microfinance. *The Grameen Bank* It is also possible for institutional lending to closely mimic and exploit some of the features of informal lending. For instance, it may be possible to design an innovative rural credit scheme in a rice-growing area in which a formal-sector institution acts both as a lender and a miller. The combined activity will permit the formal institution to accept rice output as repayment for loans. By mimicking the activities of a trader–moneylender, the formal institution may actually be able to carry out lending activities that reach small borrowers and are profitable at the same time.

Even if such mixed institutions are difficult to put into practice, there are ways in which the information base of a community can be put to use by a cleverly constructed lending institution. An example of this is the Grameen Bank of Bangladesh.[310]

In Bangladesh, as in many other developing countries, subsidized rural credit offered by government banks laid the basis for its rural credit policy. It didn't work. Repayment rates were low, and loans (or the forgiveness of past loans) were often used for political reasons, undercutting the effects of the policy.

The Grameen Bank, started in the mid-1980s by Mohammed Yunus, lends to very poor households, and lends to *groups* of borrowers rather than individuals. A typical group consists of five borrowers, and lending to individuals within the group occurs in sequence.[311] The average loan size is around \$100. No collateral is required and the

[310] Many of the observations in this section draw on Morduch [1997].

[311] In this sense the Grameen Bank also drew on the idea of Rotating Savings and Credit Associations (ROSCAS) (see Besley, Coate, and Loury [1993]), as well as credit cooperatives, from which it indirectly drew inspiration (see Woolcock [1996]).

nominal rate of interest is around 20% (roughly 12% real). Over 90% of the borrowers are women. Average repayment rates are over 97% (Morduch [1997], Christen, Rhyne, and Vogel [1994]).

Group lending and the use of information The central feature of the Grameen Bank's lending policy is that in the event of a default, *no group member is allowed to borrow again*. This means that a group has to be formed very carefully by the individuals to weed out bad borrowers who could jeopardize the creditworthiness of the group as a whole. Observe how information is being used by the Bank even though the Bank itself has no a priori access to it. The point is that the borrowers *themselves* have the incentive to use this information to form groups, and this induces a form of *self-selection* that no individual-based banking scheme can mimic.[312] Consider some of the specific implications.

Positive assortative matching. Self-selection typically takes the form of *positive assortative matching* (Becker [1981]), where good credit risks come together. To be sure, both bad and good risks equally want to team up with the good risks, but if the good risks can identify other good risks, they will form a group together. Thus group formation has the property that it can drive risky types out of the market, because some of the costs of their riskiness are borne by other borrowers instead of the bank alone. As we saw earlier in this chapter, lending to individuals has exactly the opposite effect: in the presence of limited liability, the risky types outbid the safe types (because the costs of their project failures are effectively borne by the bank). Ghatak [1996] and van Tassell [1997] develop this line of argument (see also the exercises at the end of this chapter).

Peer monitoring. The argument so far assumes that borrowers are *intrinsically* safe or risky. This is a valid assumption in many situations: group members may not be able to *control* what a fellow member does with her money, although they might still be able to predict what she is going to do (for instance, because they know her personally). In other situations, group members might be able to monitor and influence the choice of individual projects. In that case, the group will want these projects to be safe, relative to the safety levels they would choose as individual borrowers.

To see this, note once again that individual borrowing under limited liability leads to excessive risk taking, because the borrower fails to fully internalize the costs of project failure, but when borrowers congregate into groups, more of these costs are borne by fellow borrowers, who will then put pressure on one another to cut back on the level of riskiness. Thus group lending has two effects that cut in opposite directions. First, relative to individual borrowing, it *increases* the risk on any one borrower for a *given* level of project riskiness: this is a cost. Second, it creates pressures for peer monitoring to lower the level of project riskiness: this is a gain. Can the gain outweigh the loss and lead to a better institution overall? Stiglitz [1990] argued that under some conditions, it can.

[312] It should be added that the Grameen Bank is just as much a social group as a bank, It requires borrowers to observe certain codes of conduct: they make commitments to have small families, boil their water, repair and maintain their dwellings, not make or receive dowry payments, and so on. These social involvements make the Grameen Bank very different from an ordinary financial institution, and may even have spillover effects on loan repayment (as with credit cooperatives in the Philippines formed from one religious community, where Sunday prayers include pledges to repay loans to the cooperative).

Potential drawbacks. Group lending is not without potential drawbacks. Perhaps the most serious of these occurs when one member of the group runs into genuine financial difficulties and must default. In that case it is a dominant strategy for other group members to default on their outstanding loans as well, because one member's genuine difficulties have destroyed the group's credit rating *anyway* (Besley and Coate [1995]). The Grameen Bank avoids this problem by lending *sequentially* to group members, thus minimizing the contagion effect of an individual default.[313]

The peer monitoring effects might backfire as well; that is, there may be excessive pressures to choose overly safe outcomes that are not socially optimal in terms of average profitability. For instance, a group sees as its potential loss the costs of a project failure (which is a social loss as well) *as well as* the cost of future exclusion from borrowing. The latter cost necessarily exists because the no-lending threat must be used to avoid voluntary or strategic default, but this drives a wedge between the group's calculation of private cost and the true social loss of project failure (which is thereby smaller).[314] Therefore, a group will have some tendency to be *overconservative* in its choice of projects (Banerjee, Besley, and Guinanne [1992] pursue a related line of reasoning in their study of nineteenth century German credit cooperatives).

Finally, group-based schemes may lack flexibility. This can happen in rapidly changing environments in which some group members fare much better than others. In that case, the worst-performing member slows down the group as a whole (Madajewicz [1996]; Woolcock [1996]).

Viability and performance

Viability. The Grameen Bank lends to the poor. There are limits on the interest rate that it can charge without seriously affecting repayment capacity, and loan sizes are very small. This means that each loan dollar comes with a high administrative overhead. There are costs of providing, tracking, and ensuring repayment of the loan, and these are fixed costs *per borrower*, so that small loan sizes raise these costs. In addition, as we've already seen, the Grameen Bank provides several social services as well, and despite the extraordinary commitment of its staff, these are not free.

Grameen has therefore functioned under a significant amount of subsidy, both from foreign donors as well as from the Central Bank of Bangladesh. In the period 1991–92, subsidies amounted to about twenty-two cents for every dollar lent. According to calculations carried out by Morduch [1997] for the period 1987–94, the Grameen Bank charged annual interest rates that ranged between 12 and 16.6%, but it would have had to charge borrowers between 18 and 22% to cover operating costs without any direct grants. However, there are subsidies on capital costs as well, in the form of lower-than-market interest rates on loans *to* Grameen, and these are more important

[313] Sequential lending also minimizes the possibility of a coordination failure—a situation in which all the group members are solvent, but each declares default in the expectation that others will as well. The possibility of coordination failure is probably limited in small groups in which communication is easy, but can lead to self-fulfilling crises in larger groups, such as cooperatives (Huppi and Feder [1990]).

[314] I have been a little cavalier with this argument. By the standard limited liability argument, the direct costs of a project failure to the group are typically smaller than the social costs (because the bank bears some of the costs). Now adding the exclusion cost means that the net comparison is ambiguous, so overconservativism is not a necessary outcome.

than the direct grants. To cover *these* subsidies, Morduch estimated that interest rates would have to climb to 32–45%.[315]

Performance. What effect has Grameen had on the lives of people? This amounts to the empirical question, in the *absence* of the program, how much worse off would the borrowers have been? This turns out to be a very hard question to answer, precisely because Grameen has been adept at identifying good borrowers! In other words, it is not enough to look at the incomes of a borrowing household relative to the incomes of nonborrowers, even after controlling for the observable variables, such as age and education, that we (the statistical observers) can measure. The reason is that we are unable to observe characteristics such as drive, creativity, inventiveness, or entrepreneurship—traits that the borrowers themselves use when forming groups. Thus it is possible that only the more "able" borrowers have been able to access Grameen, but the incomes of such borrowers would have been relatively high even if no Grameen had existed! It follows that a regression destined to capture the effects of credit on participation will *overestimate* the positive impact of the program.

This sort of selection bias may also run the other way. Not only does Grameen look for good borrowers, it looks for poor ones: in particular, such borrowers may be poor for unobservable reasons that are not captured by the characteristics that we can measure. Now a cross-sectional regression will *underestimate* the effect of the Grameen Bank, because the incomes of those treated under the program will, on average, be lower (at least before the program). This happens because "program placement" isn't random, which is a general problem with assessment of any program that targets the underprivileged, such as setting up health services in areas with substantial undernutrition and illness, or crop extension services that are set up in areas of low know-how, or family planning programs in areas of high fertility.

One way to get around this problem is to find an *identifying variable* or *instrument*: one that affects participation in the program but does not itself affect the outcomes of participation. These are hard to find, in general. Variables that correlate with participation usually affect incomes as well, which resurrects the specter of selection bias.[316]

A potentially superior alternative is to use before–after tests on the same people. This takes care of the unobserved variation in characteristics. Of course, there may be unobserved variations over *time* that have nothing to do with the program, such as societywide economic improvements. This can be dealt with if there are *also* data on a group of people in the same locality who were not participants in the program.

[315] These calculations assume that default rates do not alter as the interest rate is hiked. If one allows for this possibility, the implied interest rates are even higher, obviously.

[316] McKernan [1996] suggests the interesting use of the number of landowning relatives of potential borrowers as an instrument. The idea is that this will affect program participation: a person with more landowning relatives is less likely to be credit constrained. Of course, the existence of such relatives might also influence income. Moreover, variables such as these have a weak influence on participation rates. Pitt and Khandker [1995] took a different approach to get around the program placement problem. Most microfinance programs have eligibility criteria that must be met for participation in the programs. In Bangladesh, eligibility is restricted to households with less than half an acre of land. Thus noneligible households can be used to get a handle on village-level effects. Now the *additional* differences at the village level among participants can be attributed to the microfinance program. The unobservability of borrower characteristics is not an issue here, because they focus on the effects of overall program availability.

Unfortunately, few microfinance programs have data *both* on before–after performance and on some "control group" that did not participate.[317]

With these effects controlled for as well as possible, it appears that the effect of Grameen is around seventeen cents to the dollar (Pitt and Khandker [1995]): that is, every dollar lent serves to raise annual household expenditure by about seventeen cents. Note that this is lower than the twenty-two cents per dollar of subsidy, and we might therefore be tempted to conclude that the program is not worth it. However, in reaching this conclusion we would have to ignore the fact that Grameen is a *targeting* organization just as much as an organization for transfers: the money seems to reach the really needy and enterprising, and this would not have been possible through indiscriminate transfers of money. You might think of the five cents on the dollar as the cost of targeting, and who's to say that this isn't money well spent?

In addition, the seventeen cents impact does not include the additional earnings that are saved (note that the amounts relate to household *expenditure*, not income) nor the value of the social education that Grameen provides. It does not include the long-term acceleration in earnings that may have been made possible, and it does not capture the possible spillover effects on nonborrowing households, although there may be negative as well as positive spillovers. Thus a definitive evaluation of the Grameen Bank's success awaits further research.

Other microfinance institutions

The Grameen Bank spawned, by example, many other microfinance programs. Morduch [1997] provides a list of the many countries that have adopted finance programs based on the Grameen model: among them are several countries in Latin America, Africa and Asia, and the United States. Some of these programs have been more successful than Grameen if one goes by the narrow criteria of economic viability. For instance, the BancoSol in Bolivia lends to groups rather than individuals, just as the Grameen Bank does, but the groups are significantly richer than the clientele of Grameen. Thus average loan sizes are larger, and so are interest rates, allowing BancoSol to turn a profit. The Bank Rakyat Indonesia (BRI) lends to individuals and asks for collateral. Again, we are talking about a clientele with economic characteristics very different from those of their Bangladeshi counterparts. Thus Grameen remains a role model on two counts: because of the pioneering and innovative inroads it has made into small-scale lending and because it has reached out to very poor individuals. This last goal has understandably pulled down its viability on the very narrow yardstick of economic profit, but how does one value overall social impact? Is it valued under the broad criteria of the last section or perhaps, using the fact that so many successful microfinance programs today owe their existence to the Grameen Bank?

Fixed versus Working Capital

20.3.3. Default and fixed-capital loans. The analysis in the earlier section is deficient in one serious respect. It assumed that the default probability is *independent* of the amount to be repaid. It is easy enough to describe situations in which this should not be the case; for instance, one of the models in Chapter 11 built exclusively on this possibility.

[317] For assessments of microfinance programs using the before–after method, see, for instance, Churchill [1995] and Mosley [1996].

Larger amounts to be repaid may lead to a greater risk of default. This statement suggests that certain loans will not be given at all under *any* circumstances, irrespective of the interest rate premium, because the *premium itself affects the chances of repayment*. Likewise, large loans themselves raise the chances of default and will, therefore, not be made. What is "large" depends, of course, on the particular circumstances of the society, such as per capita wealth and the availability of alternative opportunities such as migration to another area (in the case of default).

We can extend this line of reasoning, not just to the *size* of the loan, but also to the kind of *use* to which the loan will be put. If the loan can be used by the borrower to *permanently* put himself in a situation in which he never has to borrow again, then such loans may not be forthcoming. Suppose, for instance, that a rural laborer wishes to borrow money so that he can migrate to the city and set up a small business there. Indeed, given his contacts and entrepreneurial spirit, this may be efficient from a societal viewpoint, but it would be surprising indeed if any rural moneylender advanced him such a loan. In the absence of a legal enforcement mechanism, often the only instrument that a moneylender has is the threat of not advancing loans when needed in the future. But if future loans will never be needed, then the threat has no value.

It is therefore reasonable that in the presence of strategic default, the overwhelming provision of informal loans will be for working capital or consumption purposes, rather than for fixed investments that may permanently reduce the borrower's future need for credit.

20.4. Insurance

Insurance markets, which we will study in Chapter 21, suffer from similar sorts of complications, although repayment is not the main issue. Consider crop insurance. If such insurance is provided by a large company without intimate knowledge of the insurees, it is for the insuree to assert that the harvest was bad, with no surefire way for the insurer to verify whether this was actually the case. There are, of course, some observable shocks such as weather that the company can condition its payments upon, but very often there are farm-specific, idiosyncratic shocks as well. To the extent that these shocks are difficult to observe and verify (publicly), a company-based scheme may be unable to cover the important need for crop insurance.

Indeed, even if the *outcomes* (such as farm output) can be verified, insurance is still problematic if the *inputs* cannot be verified. For instance, there is the problem that crop insurance may encourage undersupply of effort on the farm, due to the knowledge that a bad crop will be insured against.

All of these issues are exacerbated by the fact that agricultural production also exhibits a great deal of correlation across farms. Bad weather may have an entire village or group of villages clamoring for an insurance payout. Companies that can handle such payouts must be diversified over a larger area, but such diversification does not allow for microknowledge of conditions within an individual village.

It is not surprising, then, that insurance schemes in village economies are often *informal* in nature: groups of people co-insure one another under better informational conditions. As a farmer and a village resident, it may be difficult for my next-door

neighbor to convince me that his crop has failed when I know about his activities on a daily basis. I may also be able to monitor his effort better, and he mine. However, such schemes, better grounded in information though they may be, can obviously not provide insurance against a correlated event (such as villagewide flooding). They also suffer from the second problem that we discussed for credit: enforceability. An individual who is in good shape this year may be that much less willing to help out someone who is not. He knows that in displaying this unwillingness, he may be locked out of future access to insurance (and he may face the disapproval of the community as well), but this may or may not be enough to deter him.

20.5. Basic Concepts

The main ideas that we seek to develop can be illustrated with an example that involves just two people.

Suppose that Asaf and Sharif are two farmers. We begin by supposing that they produce the same crop and that they use the same amounts of land and other inputs. Suppose that the harvest can take one of two values: $2,000 if all works out well and $1,000 if there is some damage to the crop. Assume that the chances of either of these events occurring is 1/2, and that this probability is the same for both Asaf and Sharif.

If Asaf and Sharif do not know each other, then they will try to do something about this uncertainty on their own. You have seen several examples by now of risk aversion and how this gives rise to the desire to smooth consumption, and we need not repeat the details here. Suffice it to say that both Asaf and Sharif would like to even out their consumption. They would like to run down their wealth a bit to make up for the shortfall when the crop is bad and stock some of the excess when the crop is good.

Using one's own wealth to smooth uncertain shocks in income is called *self-insurance*. Essentially, individuals with the wealth to do this are protecting themselves against uncertainties that they cannot control. Self-insurance can work through several channels. Stocks of cash or accumulated savings in banks can be run down (or added to) for the purpose. The same can be done with grain stocks, although holding such stocks can be costly because grain is not perfectly durable. However, this may be a preferred form of savings if rural banks are few and far between or if there are restrictions on the rapid liquidation of savings. Other assets may be run down or accumulated as well. Livestock and jewelry are two such assets, although these may be sold as a last resort. The box on self-insurance in Indian villages using bullocks will give you a flavor of how self-insurance works in practice.

Self-Insurance and Bullocks

Possibly the most important mechanism for consumption smoothing, other than market credit and intra- or interfamily lending, is the sale and purchase of assets. Farmers hold many different forms of wealth, including land, jewelry, currency, capital goods (such as pump sets), animals (such as bullocks), and stocks of food grain. It may be possible to use such assets as *buffer stocks*; that is, families accumulate stocks in periods of relative affluence and deplete these reserves to finance consumption expenditures in more difficult

times. Do rural households engage in substantial buying and selling of assets for this purpose? A study by Rosenzweig and Wolpin (1993), based on the ICRISAT villages in India, suggested that for certain forms of assets, this is indeed the case.

Immobile capital—land and buildings—constitute a major part of farmers' wealth in this study, accounting for approximately 85% of total wealth. However, it is a widely observed fact that although the market for land *rentals* is very active in these regions, the asset market for land is highly inactive. Of all rural households surveyed in 1970–71 by the National Council of Applied Economic Research in India, only 1.5% undertook any kind of land sale, which suggests that this method is rarely used to finance consumption in times of need. In the non-real-estate wealth category, financial assets (stocks, bonds, etc.) have a very small share (less than 5% even for large farmers), and thus play an insignificant role. Average crop inventories held over the year accounts for about a quarter of this wealth; however, it varies a lot across harvest and nonharvest seasons, and is probably used more to smooth consumption over the year rather than across years. A sizable portion (about 19% on average) of nonland wealth is held in the form of jewelry, but the data indicate that buying and selling jewelry, too, is minimal. By far the largest component of nonland wealth is bullocks (representing 50% of the wealth for small farmers, over 33% for midsize ones, and about 27% for large farmers), and evidence is strong that farmers vary their ownership of bullocks as a primary instrument to smooth consumption.

The asset and rental markets for bullocks have exactly the opposite characteristics compared to those for land. Although there is an extremely well organized, regionally integrated market for bullocks,[a] short-term bullock leases are extremely uncommon (see Chapter 22 for further discussion). The nonexistence of the market for bullock power (as opposed to the market for bullocks) implies that there can be substantial productivity gains for a farmer who owns an animal or two; thus, from a purely productivity viewpoint, turnover in bullock ownership should be expected to be low.[b] However, a ten-year survey revealed that 86% of households were involved in at least one transaction in bullocks, indicating that many of these were probably "distress sales," motivated to meet consumption requirements.

A careful analysis of the data shows that this is indeed the case. Due to the presence of well integrated bullock markets, bullock prices are largely immune to village-specific production shocks, which is precisely the factor against which farmers seek insurance. This is highlighted by the fact that over 60% of bullock sales were made to buyers outside the village, with 10% going to buyers located more than 20 kilometers away.[c] The statistical results show that the probability of a bullock purchase increases significantly when income is high, and the probability of a sale decreases. Moreover, other things held constant, those people holding higher stocks of bullocks (or even pump sets) are less likely to make a purchase in the future, suggesting that on the average, just as with any other buffer stock, farmers try to maintain a target level of the asset. This pattern is strongly borne out in data on medium and small farmers. For large farmers, the fit is much weaker, suggesting that this group has much greater access to credit and other conventional instruments and does not have to rely heavily on asset sales as a means to achieve consumption targets.

Does the widespread incidence of distress sales (coupled with the absence of rental markets for draft animals) hamper efficiency and lower average agricultural output? Rosenzweig and Wolpin felt that it does. At the prevailing market prices for bullocks and other inputs, their profit function estimates show that the optimal number of bullocks owned by midsized farmers is about two. However, the average size of the bullock stock is 0.94. Through numerical simulations of their econometric model, they estimated that the provision of actuarially fair weather insurance does not increase farmer welfare substantially. The weather is only one source of risk, and farmers already have some

insurance through limited credit. On the other hand, the creation of supplementary assured sources of income (say through a rural job creation policy) raises bullock stocks closer to optimal levels and thereby increases agricultural efficiency.[d]

[a] In India, large centralized bullock "fairs" are held in many regions at specific times of the year.

[b] Part of the potential efficiency losses arising from the absence of free flow in bullock power services is mitigated by the active rental market in land. Rosenzweig and Wolpin estimated that a farmer who did not own a bullock was 63% more likely to lease out land than a farmer who owned at least one.

[c] This observation is unlikely to hold with much generality across countries, especially in those areas where intervillage integration is relatively small. For example, Czukas, Fafchamps, and Udry [1997] test whether livestock is used as a consumption smoothing device by households in Burkina Faso. Unlike the observations for India, the authors find that the effects of random shocks on livestock sales are indeed small, suggesting both that markets are local and that there are a fair number of alternative smoothing options.

[d] We need to be a little careful, though, in interpreting these results and drawing the conclusion that a serious "underinvestment" in bullocks exists. After all, the *aggregrate* supply of bullocks must be quite inflexible at least in the short run, and can increase only to the extent that more favorable conditions lead to greater breeding. The problem is that profit and optimality calculations are performed under the *prevailing* market price for bullocks: if consumption smoothing motives are removed through the provision of alternative sources of insurance, the demand and supply conditions in this market change entirely, leading to entirely new equilibrium prices and optimality calculations. However, the mere stabilization of bullock stocks leads to some increase in output and welfare, even if the *average* size of these holdings does not increase.

The second major form of smoothing is taking recourse to credit. We have already discussed this in great detail and there is nothing that we need add here.

What I would like to focus on here is a third source of income smoothing: that attained through mutual insurance between economic agents. As the discussion continues, you will see that this source of smoothing is, after a point, difficult to distinguish from credit, but for the moment, we postpone these complications.

To develop this idea, bring Asaf and Sharif together. We can think of four possible outcomes: (1) both Asaf and Sharif produce the good output; (2) both produce the bad output, and [(3) and (4)] one produces high while the other produces low. In the first two cases there is little that Asaf and Sharif can do to help each other; they are in identical positions. In the last two cases, they can agree to share their output equally. What this boils down to is a promise that Asaf will pay Sharif $500 when he produces high and Sharif produces low, and Sharif will pay Asaf $500 when the opposite occurs.

The following observations are to be noted. First, *if the chances of producing high output are independent across Asaf and Sharif*, elementary probability tells us that the chance of each of these four cases is 1/4. Without any consumption smoothing, Asaf's consumption in the four situations is (2,000, 1,000, 2,000, 1,000); a similar statement holds for Sharif. With the mutual insurance scheme in place, this consumption stream changes to (2,000, 1,000, 1,500, 1,500). This reflects the fact that in the first two states there is nothing to be done in the way of mutual insurance, whereas in the last two cases income is equalized. Does this provide an improvement in expected utility? It certainly does, if Asaf and Sharif are risk-averse. You can see this in exactly the same way as we did in previous chapters. Mutual insurance gives rise to a mutual gain for Asaf and Sharif.

Second, observe that *at any date*, if the high–low scenario occurs, either Asaf will have to pay Sharif or Sharif will have to pay Asaf, but not both. Mutual insurance

demands that ex post, after the outcome is revealed, one party make a unilateral transfer to the other. Moreover, the slate is wiped clean after this transfer. Next year, the *same* person may be asked to pay up again. The point is that a past transfer from one person to the other does not carry a historical burden for the recipient. This is what makes a scheme of pure insurance conceptually distinct from credit. A credit scheme ties past transfers to the expected pattern of transfers in the future. An insurance scheme in its purest form does not.

Third, observe that the power of mutual insurance depends greatly on the *degree of correlation* between the outputs of Asaf and Sharif.[318] Suppose, for instance, that high or low outputs are determined entirely by the state of the weather, and not by the possibility of localized pest damage. If Asaf and Sharif are in the same village, the weather will affect them equally, and cases (3) and (4) will *never* occur. Both farmers will always be in the same position and no mutual insurance is possible. Weakening this extreme example a bit, if both outputs are highly positively correlated, there will be situations in which the high–low scenario occurs, but this will happen with small probability. There *are* gains to be had from mutual insurance, but the magnitude of the gains is small ex ante.

Reversing this argument, imagine a situation in which the fortunes of Asaf and Sharif are perfectly negatively correlated. It may be hard to conceive of such a situation in reality, but treat it as a benchmark for a moment. In this case, the high–low scenario will *always* occur and the gains from insurance are huge. Thus it appears that insurance groups are most likely to be formed among those people whose fortunes are as negatively correlated as possible. There is some truth to this statement, but there are also reasons to avoid such extremes (even if they exist), which we shall come to later.

The general point is that for mutual insurance to work, there must be a certain absence of positive correlation between the fortunes of the participating economic agents. This is one reason why crop insurance is difficult: agriculture does present large correlations because of the weather. Having said this, though, it is only fair to recognize that there are situations that *are* idiosyncratic across various economic agents (illness, festivities, local crop damage), and these can benefit from insurance along the lines described in the preceding text.

Finally, suppose that there are not just two people, but many replicas of Asaf and Sharif. Assume again that their fortunes are all independent. Now ask yourself what the *average* income of this large string of people will be in any given year. This problem is identical to a coin-tossing problem: toss a coin several times, and each time write down 1,000 if heads comes up and 2,000 if tails comes up. The average of this string will be close to 1,500 for a long string of tosses.[319] For exactly the same reason, the average income will be $1,500 (identify each independent coin toss with a person), provided that the number of farmers is large.

Now suppose each of these farmers follows the rule of contributing $500 to a common pot if their income is $2,000 and claiming $500 from the common pot if their income is $1,000. If the number of farmers is large enough, then it follows from the argument in the previous paragraph that in any one year, this rule is *viable*, in the sense that the surplus or deficit in the pot will be negligibly small. Again it is true that in any

[318] See Appendix 2 at the end of the book for a detailed discussion of this concept.

[319] This is what we know of in probability theory as the (strong) *law of large numbers*.

one year, a farmer will be a net contributor to or a net recipient from the common pot, but note that each farmer succeeds, in this way, in smoothing his income *completely*. In the two-farmer case, this would have been possible only if the two incomes were perfectly and negatively correlated, so that Asaf gets \$2,000 whenever Sharif gets \$1,000, and vice versa. However, such a state of affairs is unlikely. With a large number of farmers, independence of fortunes is enough to get full smoothing.

It is important to note that these insurance schemes are rarely constituted in some formal way. The box (to come) on reciprocal credit–insurance arrangements in Nigeria will give you some idea of how these transactions are carried out. Underlying the transactions is some social norm of reciprocity—one that pushes you to help out a community member in need. The transfer you make might be in cash or might be in kind (such as reciprocal exchanges of labor on farming plots). As we have seen on more than one occasion in this book, social norms cannot survive for long if the economic incentives that hold such norms begin to wither away, and so despite our acceptance of reciprocity and mutual insurance as a *social* institution, we are looking for the economic underpinnings that permit such an institution to survive.

20.6. The perfect insurance model

20.6.1. Theory. The example of the previous section allows us to formulate a model of perfect insurance. Suppose that a village is populated by a large number of farmers. For simplicity, let us assume that the farmers are all identical, though this is in no way necessary at a conceptual level. For each farmer, describe his income at each date by the equation

$$Y = A + \epsilon + \theta. \tag{20.5}$$

The income equation has three components. Think of the first (A) as each farmer's *average* income (this is a number such as \$1,500 in the example of the previous section). The second component (ϵ) is a random shock that may have the same distribution across farmers, but affects each one *independently*. For instance, the crop may be eaten by animals, or destroyed by bugs, or there may be inappropriate application of water or fertilizer. The shock may represent a shortfall of income due to a sudden spike in consumption needs; for instance, a part of income may be paid off for illness (in this case think of Y as net income after such necessary expenditures). It is a local or *idiosyncratic* source of uncertainty, which might happen to every farmer but happens in an independent fashion. Finally, the third component (θ) captures all the common or *aggregate* variation in the village. This is a source of uncertainty whose realization in any year affects *all* farmers in the same way. The best way to think of this is to imagine that these are common income variations stemming from the weather (the state of the monsoon, for instance).

Because we regard A as the average income, we should think of the random terms ϵ and θ as having mean value equal to zero.

Now, a little thought will convince you that if there are a large number of farmers, all the idiosyncratic variation embodied in ϵ can be insured away in exactly the same manner as in the previous section. Farmers simply pay their realized value of ϵ into a common fund. Because ϵ is sometimes positive and sometimes negative, this amounts to saying that some farmers make positive contributions, whereas others stand to receive payouts from the common fund. This way, all the idiosyncratic variation in ϵ can be ironed out. With the insurance fund in place, a typical farmer's *insured income*

is now given by

$$\bar{Y} = A + \theta. \tag{20.6}$$

Compare (21.1) and (21.2). The income stream $\bar{Y}$ carries less risk than the original stream Y, because we have removed the idiosyncratic component. Average income is still A, of course. It follows that if farmers are risk-averse, they prefer this system of mutual insurance.

What about the aggregate shock θ? Can't we pool this in the same way? The answer is no, and the reason is that the value or realization of θ is the *same* for all the farmers in the village, so there is no insurance possible here. Neglecting self-insurance and credit for the moment, the village as a whole cannot mutually insure itself against a random event that affects everybody in the same way.

Hence it is not the case that under the perfect insurance scheme insured incomes don't fluctuate at all: they do, but the source of fluctuation is the aggregate uncertainty in the system that cannot be insured away with the use of a common pool of funds. This line of reasoning also suggests that one important determinant of mutual insurance is the relative significance of idiosyncratic to aggregate risk.

20.6.2. Testing the theory. The model that we described also yields a natural test for the existence of perfect insurance in a community such as an entire village. Controlling for movements in *aggregate* village consumption (which act as an indicator of aggregate uninsurable risk), fluctuations in individual household income, or, indeed, idiosyncratic shocks such as unemployment should *not* have any effect on the consumption of that household.

As we have seen on more than one occasion, regression analysis is a good tool for such a test.[320] Use the average consumption of the community and household income as independent variables. If possible, also include consumption-side shocks (such as unemployment) as independent variables. The dependent variable is household consumption. Roughly speaking, if the theory of perfect insurance is right, the estimated coefficients on all the household-specific variables should be zero, and that on group average consumption should be unity. Individual consumption should track group consumption more closely than it tracks individual income.

Townsend [1993] explored this implication using three villages from the ICRISAT study, which should by now be familiar to you from repeated references in the last few chapters. Recall that the ICRISAT villages, located as they are in the semi-arid tropics of southern India, all display a high degree of risk. These risks stem primarily from crop cultivation, but spill over to labor markets, so that a landless laborer who earns wages faces a substantial amount of risk as well. The three villages studied were Aurapalle, Shirapur, and Kanzara. Household consumption data are available annually for a sample of forty households in each of these three villages for the period 1975–84. There is substantial variation in household income, and if this variation is taken as a measure of household risk, then idiosyncratic components are very much in evidence. The best way to see this is to compare the *differences* between household income and the average income of the sample, for different households. Subtracting the sample average is a bit like removing the aggregate risk. Then, if we interpret what's left as idiosyncratic risk, it is certainly the case that there is a lot of it: this residual appears

[320] See Appendix 2 at the end of the book for an introduction to regressions.

to fluctuate independently across households (see Townsend [1993, Figure 1]). This means that the three ICRISAT villages present a potentially fertile case for testing the perfect insurance hypothesis.[321]

A major problem arises, however, in the construction of an appropriate regression equation. We have household-level consumption, and measures of household income. What do we use for average consumption? The answer to this question depends on whom we suspect to be the relevant group that is engaged in mutual insurance. One possibility is that the *entire* village is engaged in mutual insurance, as one group so to speak. In that case we include some measure of village average consumption. However, all we have are data from the sample!

One option is to use the sample consumption average as a proxy for the village average and hope that by the law of large numbers this gives us a good estimate of the village average. However, this runs us into another sort of problem: even if we perform the regressions household by household, if the sample is large, the *average* estimated coefficient on group consumption must be unity! This is not a sign that the perfect insurance model passes with flying colors, but just that individual consumptions in the sample must sum to the sample average (Deaton [1990]).

One way to get around this problem is to run the regression household by household (using the sample average exclusive of the household in question) and see whether *each* of the estimated coefficients on group average (and not just the average of these coefficients) is close to a value of 1. The results of this exercise are mixed.[322] Depending on the length of the time period and the choice of consumption categories in the regressions (see footnote 339), the coefficients are sometimes, but not always, close to unity. A formal statistical test of the hypothesis that the coefficients are indeed 1, is not rejected using the full set of data, but the power of this test deteriorates with other variants of the regression.[323]

The observation that the coefficients on household income are indeed close to zero is more robust and suggests that variations in household income do not significantly affect household consumption. This outcome suggests that quite a bit of smoothing is taking place, but whether it is due to insurance per se or some other form of smoothing (such as self-insurance or credit) is difficult to say.

Another way to avoid the problem of missing aggregate data on consumption is to consider *differences* between household consumption and the sample average of households. This is tantamount to subtracting group averages from both sides of the regression, equation, so that the right-hand side now contains only individual terms, such as household income of various categories, and dummy variables for sickness or unemployment. Now, to test whether group consumption moves individual

[321] At the same time, be aware that there is substantial measurement error in estimating household incomes. This sort of error invariably biases estimated coefficients on household income variables toward zero. Some of this error can be avoided by averaging the income data over households with the same crop/plot technologies, but this remains a serious issue. Measurement error can also be allowed for (to some extent) by statistical correction. See Townsend [1993, Section 6] for more details.

[322] A particularly nasty problem is that data collection between the years 1982–84 dropped some consumption categories. To be sure, this has the spurious effect of making individual consumption move with household consumption. To sidestep this problem, we can use a restricted set of consumption categories throughout (such as grain consumption) or perform the regression by leaving out these last two years. The results concerning the comovement of household and group consumptions are predictably weaker.

[323] For a discussion of hypothesis testing, see Appendix 2 at the end of the book.

consumption more than individual income does is simply a matter of whether or not the coefficients on the individual variables are zero. In contrast to the earlier regressions, these coefficients are now significantly different from zero, but their magnitudes are small.[324] For instance, the effect of household income on individual consumption seems to be no more than 14%, which is pretty small. This is like saying that a one dollar increase in contemporary household income increases consumption by no more than fourteen cents. Moreover, events such as sickness (leading to loss of work) are insignificant in the determination of household consumption.

The preceding findings suggest that although we cannot be sure that mutual insurance is at work here, there is substantial smoothing accomplished by ICRISAT households through some means or other. Informal credit markets probably account for a large chunk of this. There is separate evidence that ICRISAT households smooth fluctuating income through credit transactions (see Cain [1981] and Walker et al. [1983]). Other mechanisms might include grain inventories (Paxson and Chaudhuri [1994]), holdings of cash (Lim and Townsend [1994]), and even the sale and purchase of livestock (Rosenzweig and Wolpin [1993] and the box on self-insurance and bullocks).

Note, however, that among the growing number of empirical studies conducted on this question, the ICRISAT villages do relatively well in exhibiting a great deal of consumption smoothing. Townsend [1995] reports on similar regressions carried out on Thai data. The results seem to convincingly reject the presumption of perfect risk sharing, although in other studies on Thailand, a significant degree of smoothing cannot be ruled out (Paxson [1992] used data on Thai rice farmers to make this point). Similarly, Deaton's [1994] study of the Côte d'Ivoire suggests that although there is comovement of household consumption, full insurance can be convincingly rejected. Household income does significantly influence household consumption. Townsend [1995] provides a discussion of these and other studies.

Even among the ICRISAT villages, the ability to smooth may vary significantly across households. Using a somewhat different test for insurance than the one described here, Morduch [1995] found substantial evidence of consumption smoothing among relatively better-off farmers, but not so for small farmers and landless laborers (Townsend's work contains similar observations as well). Morduch also made the point that insurance might look good simply because households try to smooth their *income* streams so that the remaining fluctuations can be absorbed by the available consumption smoothing mechanisms. For instance, some households may forego the cultivation of a crop with high expected return simply because the yield is more risky and cannot be smoothed through insurance. In this case, the household chooses a "safer" crop and its income fluctuations are therefore smaller and consequently easier to smooth.[325] The point is that there is a hidden cost (which is the expected value sacrified by choosing a safer crop). Thus, Morduch [1995, p. 108] concludes, "for less well off households in the ICRISAT villages, production choices will be made with an eye to reducing the likelihood that the shocks will happen in the first place."

[324] See Appendix 2 at the end of the book for a detailed discussion of the difference between *significance* and *magnitude*.

[325] The same sort of observation is true for the *way* in which a given crop is produced. Rosenzweig and Binswanger [1993] investigated this for the ICRISAT villages; see also the earlier study by Bliss and Stern [1982] on fertilizer choice in the northern Indian village of Palanpur.

20.7. Limits to insurance: Information

The general failure of the perfect insurance model to fit the facts suggests that there are limits to the ability of households to insure one another, even if the income shocks are idiosyncratic. Where do these limits come from? The answer is based on the same principles that informed our study of credit markets: limited information and limited enforceability.

20.7.1. Limited information about the final outcome. Think of two sorts of informational problems. One is that a person may ask for or expect transfers from his community while providing them with deliberately wrong or misleading information regarding his economic state: he might lie about the size of his harvest, he might lie about illness in the family, or, at the very least, he might be in a *position* to lie, in the sense that his community members do not have information to verify that the claimed occurrences indeed occurred.

Now this is certainly a possibility if we extravagantly expand the notion of what a community means. It is unlikely (even if speedy electronic transfers of funds were possible between villages) that a farmer in one extreme of the state of Bahia would be engaged in insurance with another unknown farmer in the other extreme of that state.[326] Even if such insurance were mutually beneficial, it is very difficult to verify the underlying circumstances on which that insurance must be based. So it's certainly true that this sort of informational barrier kicks in and precludes insurance over very large anonymous groups or spatial distance, even if such insurance could, in principle, be profitable for all concerned.

However, does this sort of informational problem occur within the village? It is hard to say, and the answer must surely vary across villages. One of the great strengths of traditional society is that they are highly endowed with *social capital*. Much of social capital is simply information flow. *A* knows if *B*'s harvest has been bad or if *B*'s father-in-law really has food poisoning. Social capital provides a fund of socially available information that permits a community to interact in ways that would not be possible in an anonymous society. Insurance is one such way, and there are others. We already discussed social capital in the context of credit in Section **??** and earlier in the context of permanent labor markets in Chapter 23. In villages where mobility is on the increase, this sort of knowledge begins to die away, reducing the scope of insurance networks. Likewise, a village that is segmented into subgroups (say, along caste lines or perhaps even along economic hierarchies such as landowners and the landless) may be unable to develop insurance schemes that cut across these lines of segmentation, simply because the flow of information is restricted. All of these theories suggest the need to extend empirical work to define more precisely what we mean by the community, and to the extent that our notion of community fails to capture these finer points (mainly because of lack of data), we might fail to capture the existence of insurance "subnetworks" simply because we are looking for one giant network, which doesn't exist.

20.7.2. Limited information about what led to the outcome. The second kind of informational problem is more subtle and, as we will see, does not altogether rule

[326] This does not prevent organized financial institutions from indirectly linking them, but that is another matter.

out the possibility of insurance. This is the problem of *moral hazard*—the possibility that the occurrence of some insurable event can be influenced by the unobservable actions of the individual. We've already encountered moral hazard in our discussion of land markets (Chapter 22) and in our discussion of credit rationing (Chapter 19). Clearly, moral hazard matters for insurance as well.

Consider mutual insurance of harvests. Everyone knows that there are idiosyncratic shocks that can affect the state of the harvest on a particular plot. Water, fertilizer, or pesticides may have been badly applied (inadvertently), or there may be damage to the crop caused by events outside the farmer's control, or (and this is the source of the problem), the crop may be bad because the farmer *deliberately* skimped on the use of these inputs. The size of the harvest may be visible for all to see (so that the first informational problem is irrelevant), but *why* the harvest is what it is requires information of a different kind altogether. The problem, of course, is that in the presence of full insurance, the incentive to deliberately underapply inputs is higher.

To see this in its simplest form, consider a community of a large number of identical farmers. Assume that all uncertainty is idiosyncratic. Just as in the example of Asaf and Sharif, we suppose that there are two possible outputs, which we denote by L (for "low") and H (for "high"). The twist in the story is that for each of the farmers there are *two* possible probabilities of producing the high output: call them p and q, and suppose that $p > q$. The idea is that with greater care and better application of inputs, each farmer can produce H with probability p, but if he underapplies inputs, the probability falls to q. Denote the extra cost of diligent effort and input use by C (this is expressed in utility units). Finally, let u be the utility function that describes how net outputs are translated into units of utility for each of the farmers.

Begin by analyzing the farmer's problem without insurance. If he diligently farms the land, his expected utility is $pu(H) + (1-p)u(L)$, but he incurs the extra cost C up front, so his *net* expected utility is given by the expression

$$pu(H) + (1-p)u(L) - C.$$

On the other hand if he does not apply effort, the probability of the high output falls to q, although he saves on the extra cost, so that his expected utility is now

$$qu(H) + (1-q)u(L).$$

Assume that in isolation, each farmer would like to put in the extra effort; that the probability gain of the high output makes it worthwhile to do so. This amounts to assuming that

$$pu(H) + (1-p)u(L) - C > qu(H) + (1-q)u(L). \tag{20.7}$$

Now with insurance we can do even better. We already saw in several examples that if the farmer is risk-averse, then the utility of the expected output exceeds the expected utility of output,

$$u(pH + (1-p)L) > pu(H) + (1-p)u(L),$$

so there are clearly gains to be had from insurance. Moreover, with a large number of farmers, it is possible to set up an insurance scheme that (approximately) guarantees this for every farmer. To see how, suppose that every farmer with high output pays the sum $(1-p)(H-L)$ into a common pool, while every farmer who produces the low output receives a payout of $p(H-L)$. Do the necessary simple algebra to verify

that with this scheme, each farmer gets the average value of output, $pH + (1-p)L$, *for sure*. Now all we have to do is to make sure that the scheme is feasible: that total payouts balance total receipts. To see this, note that with a large number of farmers, approximately a fraction p of them get the high output, so that the total inflow into the pool is (approximately) $p(1-p)(H-L)$. At the same time, approximately $(1-p)$ of the farmers produce the low output, so that the total outflow from the pool is $(1-p)p(H-L)$. Inflows and outflows match, so we are in business.

However, there is a flaw in this argument. We *are* in business if we can somehow guarantee that each farmer continues to put in the high level of effort. But it is not possible to do so under this insurance scheme if the extra effort cannot be verified by outside parties. If each farmer receives $pH + (1-p)L$ *regardless* of whether his output is high or low, then why incur the extra effort to raise the probability that output will be high? Thus, each farmer has an incentive to slack off under the perfect insurance scheme, and, of course, if all farmers think this way, the scheme cannot survive.

Notice that the problem is the same as that encountered in the principal–agent model. There is a trade-off between the provision of insurance and the provision of incentives. If the community wants to maintain effort at the higher level, it will have to offer incomplete insurance so that individual incentives are not destroyed. How much insurance can it safely offer? To solve this problem, suppose that X is a farmer's net consumption when output is high and Y is his net consumption when output is low, in any insurance scheme. The requirement that the scheme be approximately viable is equivalent to the equation

$$pX + (1-p)Y = pH + (1-p)L. \tag{20.8}$$

If X and Y are too close together, people will slack off. To see how close they can be, take a second look at equation (21.3), and substitute X for H and Y for L. Note that high effort will be chosen as long as

$$pu(X) + (1-p)u(Y) - C \geqslant qu(X) + (1-q)u(Y)$$

or equivalently, rearranging these terms, as long as

$$(p-q)[u(X) - u(Y)] \geqslant C.$$

Thus the *closest* that X and Y can get without upsetting the desired incentives is given by the equation

$$(p-q)[u(X) - u(Y)] = C. \tag{20.9}$$

Equations (21.4) and (21.5) are two equations in two unknowns. Use them to solve for X and Y. This is the *second-best insurance scheme*—the best one can do when incentives pose a constraint.

Note that in any solution, it must be the case that $X > Y$ [simply examine (15.5)]. This means that under the second-best scheme, *individual consumption must move with individual income*. Such a finding is consistent with the data, although to empirically test this model against others requires finer analysis.[327]

To complete this model, consider a numerical example just to see how much insurance is restricted by these considerations. Let $H = 2{,}000$ and $L = 1{,}000$ just as in the Asaf–Sharif example. Let $p = 3/4$ and $q = 1/4$. Suppose that the utility function is given by $u(x) = \sqrt{x}$. Finally, let the additional cost of diligence assume

[327] Ligon [1993] pursues this line of research.

the value 2 in utility units. Check that the condition (21.3) is comfortably satisfied, so that in isolation each farmer will indeed put in the higher level of effort. Now we solve for the second-best insurance scheme, described by the pair of numbers X and Y that solve (21.4) and (21.5). These two equations reduce to

$$\frac{3}{4}X + \frac{1}{4}Y = 1{,}750$$

and

$$\frac{1}{2}[\sqrt{X} - \sqrt{Y}] = 2.$$

Solving, we see that X is approximately \$1,832 and Y is approximately \$1,505. Thus farmers with good harvests pay out around \$170 each and farmers with bad harvests receive around \$500 each. More insurance would certainly be desirable but cannot be achieved, because the constraints on incentives would be violated.

Informational constraints pose a real problem to effective insurance. Groups with better access to the information of their members are generally in a better position to provide mutual insurance. This is also true of groups that are composed of members who feel some altruism toward one another. In this case, the incentive constraints are somewhat loosened, because each member internalizes (perhaps partially) the undesirable implications of his deviation for other group members. Hence, it will not be surprising to see that insurance arrangements between members of extended families or of the same kin group[328] are relatively common.

The flip side of this argument is that diversification possibilities within the extended family may be relatively limited. This will be the case if all members participate in a similar activity and are therefore exposed to the same set of aggregate shocks. We expect, then, that families will attempt to diversify their activities or spatial locations to lower the correlation between the incomes of family members. Rosenzweig and Stark (1989) found a tendency for marriage and kinship ties to be established across families from different geographical regions, perhaps to maximize insurance possibilities while retaining (or creating) the altruistic ties that are needed to counterbalance the absence of information. In the same spirit, Paulsen [1995] studied remittances from migrants in cities to their families in the Thai countryside.

20.8. Limits to insurance: Enforcement

We now turn to a different limit on insurance, that has to do with the *enforcement* of insurance arrangements.[329] Recall that mutual insurance schemes are rarely contracts that are written down in black and white and routinely backed or enforced by the law. Rather they are informal arrangements, set in the context of a social norm that encourages reciprocity and punishes deviations through social sanctions such as ostracization or public rebuke. This, too, is a form of *social* capital.

Whether a village's endowment of social capital is enough to permit a system of mutual insurance to function is a question that can only be settled through detailed

[328] For studies of reciprocal arrangements within kin groups, see La Ferrara [1997].

[329] For a more formal treatment of some of the issues in this section, see Coate and Ravallion [1993], Fafchamps [1996], Kletzer and Wright [1995], and Thomas and Worrall [1994].

empirical work, but it is useful, anyway, to discuss the various issues that arise with the help of a simple model.

20.8.1. Enforcement-based limits to perfect insurance. Suppose, once again, that there are a large number of farmers. The simplest case is one in which all uncertainty is idiosyncratic, so let's begin with it. Assume that there are two outputs, H and L, and that the high output is produced with probability p. Because we have already discussed the moral hazard problem, we take the probability of the high output to be fixed in what follows (equivalently, effort levels are perfectly monitored).

Consider perfect insurance. We already saw what this gives each person. With a large number of farmers, each farmer receives the constant amount $pH + (1-p)L$ regardless of what he produces. Let this "middle amount" be denoted by M. We already discussed the possibility that such a scheme can be undermined because of moral hazard, so this is not an issue we consider here.

Now picture in your mind a situation where this scheme is in place and transfers are made according to it year after year. Consider a particular year in which a particular farmer produces a high output. In that year he *could* receive a utility of $u(H)$ if he pocketed this output in its entirety, but the insurance scheme dictates that he give up some amount to the common pool, so that his utility under the scheme is really $u(M)$, which is obviously smaller. We can use this information to construct a measure of the *gain* enjoyed by a high-output individual if he decides to ignore an existing insurance arrangement and simply consume the output himself. Denoting the gain by G, we have

$$G = u(H) - u(M).$$

This is a measure of the incentive *not* to conform to the arrangement. Of course, this is only one side of the story. If the farmer deviates in this fashion, he will presumably be excluded from future access to insurance.[330] There is also the question of social sanctions, as we already discussed. All of this adds up to a future *loss* if he fails to conform today.

Just as we noticed for credit markets (Chapter 14) and (permanent) labor markets (Chapter 13), the size of the loss from future exclusion depends on the farmer's farsightedness. In each future date he loses the amount $u(M) - [pu(H) + (1-p)u(L)]$, which is just the difference between his insured utility and his stand-alone expected utility. If he has a mental horizon of N periods, the total loss is just this expression multiplied by N, the number of dates for which he thinks ahead. To this add a loss S, which is a measure of the utility damage imposed by social sanctions (such as a public scolding in the main village square). The total loss (L) from a deviation is thus given as

$$L = N\{u(M) - [pu(H) + (1-p)u(L)]\} + S.$$

Now put together the gains and losses to check if perfect insurance is indeed viable. For viability, we must have $L \geqslant G$, and this gives rise to the fundamental *enforcement constraint*

$$N\{u(M) - [pu(H) + (1-p)u(L)]\} + S \geqslant u(H) - u(M). \tag{20.10}$$

[330] Informational considerations are once again relevant here. Depending on how easily the news of his "crime" spreads through the community, this individual may find it more or less easy to become part of another group. We already discussed such issues in the context of credit and labor markets. They are probably less relevant in the case of insurance, involving as it does a larger group of people at a time.

This looks like an ugly equation (and it is), but it gives us some insight into the nature of the enforcement constraint.

First, look at the role of S, which is the punitive power of social sanctions. Clearly, the larger is S, the more likely it is that (21.6) holds. Second, the enforcement constraint is also more likely to hold if people value future interactions to a greater degree. This notion is reflected in a larger N. In part, a larger value of N may be interpreted as greater foresight or greater patience on the part of the farmer, but it can also be interpreted as a measure of greater faith in the durability of the scheme itself. If mutual insurance is expected to continue in the indefinite future, the loss from a single deviation is perceived to be large, and so the deviation is less likely to occur. This also brings out a "bootstrap" characteristic of such schemes (which appears in other situations as well, such as cooperative credit). The perceived instability of the insurance scheme may be self-fulfilling: N is driven down, (21.6) fails, and the scheme falls apart simply because it is expected to. Social institutions are very much like the outcomes of coordination games, and multiple equilibria are surely possible.

Third, is it the case that the greater the scope of insurance the more likely it is that the scheme will survive? This sounds pretty obvious but actually (21.6) tells us that it isn't generally true. Insurance is more beneficial when stand-alone outputs are more dispersed; in this example, when there are large differences between L and H. To be sure, this raises the perceived losses from not participating in the scheme, but it also means that more has to be contributed when the output is good [recall that the contribution is $(1-p)(H-L)$, which increases with dispersion]. This means that there is more to be gained from a deviation. The net effect can go either way.

20.8.2. Enforcement and imperfect insurance. What if equation (21.6) fails to hold? This means that *perfect* insurance is unattainable. We are then in a situation analogous to that discussed in the case of informational failure, where, nevertheless, some more limited form of insurance was possible. We can follow in the steps of that analysis to see what happens here. Once again, denote by X the consumption of a high-output farmer and by Y the consumption of a low-output farmer. For the scheme to be viable, the "insurance budget" must balance; this is just equation (21.4) which we reproduce here:

$$pX + (1-p)Y = pH + (1-p)L = M. \tag{20.11}$$

Next, the enforcement constraint must be satisfied. This amounts to the statement that the gain from a deviation must not outweigh the future losses. The one-time gain is $u(H) - u(X)$. The loss is

$$N\{[pu(X) + (1-p)u(Y)] - [pu(H) + (1-p)u(L)]\} + S,$$

so the enforcement constraint may be rewritten as

$$\begin{aligned} N\{[pu(X) + (1-p)u(Y)] - [pu(H) + (1-p)u(L)]\} \\ +S \geqslant u(H) - u(X). \end{aligned} \tag{20.12}$$

It will help to rewrite this in a way so that the Xs and Ys are all on one side. Transposing terms, we obtain the equivalent condition

$$\begin{aligned} u(X) + N\{pu(X) + (1-p)u(Y)\} \\ \geqslant u(H) + N\{pu(H) + (1-p)u(L)\} - S. \end{aligned} \tag{20.13}$$

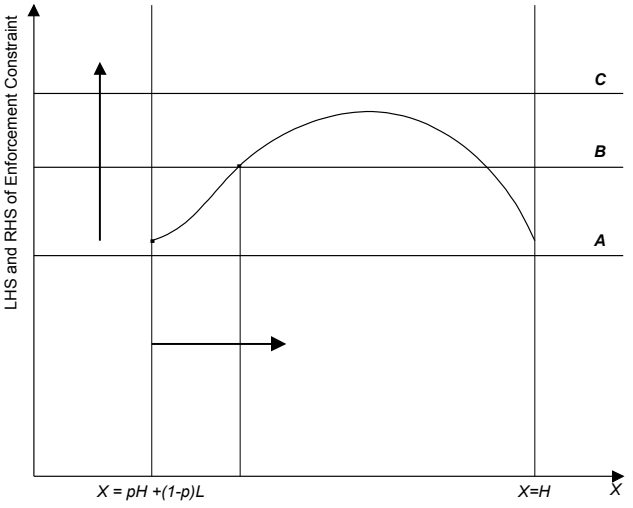

Figure 20.4. The enforcement constraint and imperfect insurance.

We are after the second-best insurance scheme. We already know that everybody wants as much consumption smoothing as possible (subject to the enforcement constraint), so we are looking for the "least dispersed" values of X and Y that are compatible with both (21.7) and (21.9). There is an easy way to do this graphically. Figure 21.1 plots how the left-hand side of the enforcement constraint (21.9) changes as we alter X, starting from the perfect insurance level [which is, of course, $pH + (1-p)L$] and moving up to the no-insurance level H. Implicitly, the value of Y is altered, so that the viability condition (21.7) holds throughout this exercise. Thus Y *falls* from the perfect insurance level [also $pH + (1-p)L$] to the no-insurance level L (but this change is not explicitly carried in the diagram).

What happens to the left-hand side of (21.9)? *At* the perfect insurance point $X = pH + (1-p)L$, the value of this expression must rise. The reason is that the term $pu(X) + (1-p)u(Y)$ is approximately unchanging for small variations in X, at the perfect insurance point (because it attains a maximum there).[331] However, $u(X)$ increases with X, so the combined expression $u(X) + N\{pu(X) + (1-p)u(Y)\}$ must rise with X.

What does this mean in words? It means that if we move away from the first-best insurance level, enforcement becomes easier. This is because the deviation gain falls, while the losses from future denial of insurance do not fall as fast, at least initially.

After a while, this trend reverses itself. The fall in $pu(X) + (1-p)u(Y)$ starts to become significant for higher values of X, and ultimately this dominates the gain in $u(X)$, so the combined expression reaches a maximum and then turns down. It is possible that by this time, X has already reached the no-insurance point H so that, effectively, there is no turning down of the combined expression. Which particular case occurs depends on the parameters of the problem and is of no great significance to us, so we also put in a declining segment in Figure 21.1 for completeness.

[331] This and the observations to follow can be easily verified by taking derivatives of the relevant expression with respect to X, accounting at the same time for the comovement of Y. I omit the details.

The horizontal line in Figure 21.1 represents the right-hand side of the enforcement constraint (21.9). It is flat because it is independent of the value of X. Note that its height depends on various parameters, among them the value of social sanctions. The lower the line, the greater the range of insurance schemes that satisfy the enforcement constraint.

We draw three instances of this horizontal segment. For the values represented by line *A*, perfect insurance is enforcable and there is no need to look at other schemes. Now imagine that social norms in this economy begin to wither away, so that the value of social sanctions declines. This has the effect of shifting the horizontal line *upward* to the line marked *B*. At this level, perfect insurance falls apart. The second-best insurance scheme is achieved by a higher value of X (and consequently a lower value of Y). This means that individual consumption has to move with individual income. The lower the value of social sanctions, the more will be this variability. Finally, there comes a point (shown by the line marked *C*) at which *no* insurance is possible at all.

Thus the lack of enforceability limits insurance and creates second-best insurance structures in which people must face, to some extent, the variation in their own individual fates. This model may be modified in many different ways, to yield different insights.

Self-insurance and mutual insurance First, we can interact self-insurance or credit with mutual insurance. Following a deviation from mutual insurance, it may be possible to self-insure to some extent. This means that punishments are effectively lower, and the tendency for mutual insurance schemes to fall apart is thereby accentuated. The greater the possibilities for self-insurance, the more likely it is that mutual insurance schemes will begin to crumble. At one level this is only natural: as alternatives appear, informal substitutes can die out without much difficulty. But it is important to appreciate that the process exhibits a built-in inefficiency. The value of a deviation from mutual insurance is the *current* (high-output) return plus the future return from self-insurance. If this value is larger than the returns from mutual insurance and if there are no additional social sanctions, the mutual scheme will collapse. However, note that *at* the collapse point, the mutual scheme generates larger value than self-insurance does, provided that the power of social sanctions is small.[332] In this sense the transition is inefficient.

Correlated shocks We also can extend the model to include aggregate (correlated) shocks to output (see Coate and Ravallion [1993]). It is easy enough to write down an enforcement constraint for this problem. Such a constraint has an interesting feature. If the entire economy is in an aggregate depressed state, the farmers who do *relatively* well in that state will have a greater incentive to deviate from any given mutual insurance scheme. This follows from diminishing marginal utility: when total output is lower, the value of *not* making some given contribution to an insurance fund must rise. At the same time, future expectations of gains from insurance are unchanged (if the shocks are independently distributed over time).

Consequently, it becomes more difficult to satisfy the enforcement constraint when times are bad for the economy as a whole, and so insurance weakens in such times.

[332] The current payoff from a deviation, if a farmer has produced the high output, is always larger than the return that any insurance scheme can provide. It follows that if S is small, the mutual insurance scheme fails to be enforceable at a point when it still does better than self-insurance. You may want to formally prove this by first adapting the enforcement constraint (21.6) to include a weaker self-insurance scheme following the deviation. Put $S = 0$. Now examine the point at which mutual insurance fails the enforcement constraint.

However, we must be careful: contrast this phenomenon with one in which there is a permanent improvement in the lot of one (or some) members of the group. Faced with a permanent improvement, they too will not contribute to mutual insurance. Thus an aggregate state of economic depression or a selective state of well-being can both contribute to the destruction of enforceable insurance opportunities.

The blurring of credit and insurance Our third (and last) extension is much more subtle than the other two and has profound consequences for the conceptual distinction between credit and insurance. Recall that the distinguishing feature of insurance is a certain lack of regard for history. It is possible (though perhaps unlikely) that Asaf may have a bad output three years in a row, and Sharif may have enjoyed a high output for those three years. Our insurance scheme dictates that Sharif must make contributions to Asaf for those three years. Moreover, once the three years are over, the slate is wiped clean again. Thus pure insurance has a unilateral air to it: transfers do not carry an obligation to repay. Indeed, strange though it may sound, this is a necessary property of the perfect insurance scheme in the model that we described. It's not that such schemes are entirely missing in the world we live in: for instance, medical insurance has that property—but it is harder to imagine informal schemes with this property.

It turns out that the enforcement constraint throws some interesting light on this issue. It is true, as we already mentioned, that the first-best scheme has a "history-independence" property. However, what if enforcement constraints prevent implementation of the first-best scheme? This is precisely the situation that we have considered so far, yet in formulating our second-best scheme, we stuck firmly to pure insurance and did not allow past transfers to influence the future. It turns out that if we drop this requirement, we can do even better (Fafchamps [1996]; Kletzer and Wright [1995]; Thomas and Worrall [1994]).

To see this better scheme, return to the formulation of the second-best arrangement pursued in equations (21.7)–(21.9) and Figure 21.1. Recall the basic idea in this construction: high-output individuals may be reluctant to contribute to the insurance scheme because they stand to gain by keeping their entire output. They balance these gains against the losses of not being able to participate in the scheme in the future (there are also social sanctions, but we neglect these here as they are not needed to make the main point). The way we previously solved for X—the consumption of the high-output person—was by balancing these two conflicting forces in a way that was consonant with the maximum possible degree of smoothing. In doing so, we neglected another source of balance; that is, *the scheme could be designed to increase, for a high-output person, the future returns from participation in the insurance scheme*. This raises the losses from future nonparticipation and thus loosens the enforcement constraint and permits a greater degree of smoothing today.

This means that the production of high output today raises a person's share of the pie in tomorrow's scheme and likewise lowers the portion for a low-output person who receives a handout from the scheme today. However, such an adjustment doesn't affect the low-output farmer's enforcement constraint: he is *receiving* money from the system, so there is no constraint to worry about in this case.[333]

[333] This is true provided we are careful not to reduce his participation utility from tomorrow onward *below* the level he can attain on his own.

Of course, this state of affairs is not permanent: today's high-output person may be tomorrow's low-output producer, and then he will slip down a notch as far as the subsequent share in the insurance fund is concerned. Likewise, today's low-output producer may recover his "status" by producing high in the future.

The algebra of this formulation is complicated and we will not pursue the details here, but you should be able to see how the basic idea works, and—more importantly—its implications. By linking an individual's payouts in the past to current claims in the insurance scheme, the scheme ceases to be one of pure insurance and acquires some of the characteristics of credit. Your current entitlement is higher if you have paid out more in the past, which sounds, to some extent, like you are entitled to a "repayment" of some of the money that you contributed yesterday. However, this is an entitlement only *on average*. It may be that your compatriot does badly again, and then the insurance aspect of the scheme will do battle with the credit aspect: you may be "required" to forgive or roll over some of the past "debt" and even advance further "loans," so it's more complicated than pure credit. Still more interesting is the interpretation of repayment from your compatriot depending on *your* economic state today: he will be required to pay you more if your current economic state is bad, relative to when it is good.

This seemingly esoteric mixed-credit insurance scheme brings us to the frontiers of current research in this area. Is it all esoterica in reality? The answer, surprisingly enough, is no: as empirical investigations unfold, we see that a great deal of informal financial interaction in developing countries is characterized by norms of both reciprocity and obligation. The former demands repeated support if circumstances require it, so it has the flavor of insurance. The latter demands some accounting of the past, so it has the flavor of credit. We end this chapter with a box that describes a pioneering empirical study in this area.

Credit as Insurance: A Case Study from Northern Nigeria

How can markets and economic institutions adapt to missing information and problems of enforcement? We have seen how these constraints can limit reciprocal transactions and force a blend of credit and insurance to develop. A study of rural credit markets in northern Nigeria by Udry [1993, 1994] provides a rich empirical illustration of institutional adaptation.

Nigeria is substantially Islamic: around half the population is Muslim. Islamic law strongly condemns the practice of charging interest for loans. Therefore, usury has always been taboo in traditional Islamic societies. Fixed repayment periods are also prohibited: "And if the debtor is in difficulty, then (there should be) postponement to a time of ease" says the *Koran*. (2: 280; quoted in Udry [1993]). At the same time, Nigerian agriculture is subject to seasonal fluctuations and random shocks (droughts, crop disease, etc.). Thus income flows are naturally highly uncertain and irregular. The need to borrow and lend in times of distress is therefore paramount. Given the stricture against charging interest, which should strongly dampen the incentive to lend, how do economic institutions deal with the problem of consumption smoothing? Is there any significant credit market alive in rural Nigeria? Udry's study, based on data collected from four randomly selected villages near the city of Zaria in the Kaduna state of northern Nigeria, shows that the answer is very much in the affirmative and that credit implicitly takes the form of interhousehold insurance against income shortfalls.

Of the 400 sample households, only 10% reported *not* having engaged in any borrowing or lending activity over the period of the survey. Indeed, 50% of households *both* borrowed and lent at various points during this interval. A noticeable feature of these loans is that they were almost always restricted within tightly knit communities, whose members knew about each other's conditions, needs, and activities. Ninety-seven percent (by value) of the 808 informal-sector loans recorded were either between residents of the same village or between relatives. Almost all such loans were remarkably informal in character: there was no written contract or witness, nor was a rate of interest or repayment date specified, implicitly or explicitly. Loans seem to be advanced purely on the basis of trust.

What is interesting about the Nigerian credit market is that the amount of money paid back depends not only on the borrower's financial situation at the time of repayment, *but also on that of the lender*. Udry's econometric analysis confirmed that the importance of the latter factor in determining the implicit rate of interest is statistically significant. This led him to conclude that the loans were not, in fact, equivalent to "equity shares" in the borrower's farm or business, but mixed devices with components of insurance and mutual risk pooling embedded in them.

It is not surprising, therefore, that the implicit realized rate of interest, calculated from loaned and repaid amounts, showed large variations; however, the median nominal monthly rate of return was about zero. Given that the monthly rate of inflation was a steep 3.7%, the average loan clearly earned a significantly negative real rate of return. This further strengthens the interpretation that the loans were really forms of mutual insurance, based on need and reciprocity. However, they were not unilateral transfers either, as a pure insurance would dictate. A transfer today carries *some* obligation to repay tomorrow.

Theory suggests that a mixed credit–insurance contract is symptomatic of enforcement problems. A household that receives a loan may decline to repay later or may repay only a meager amount. Likewise, households experiencing relative prosperity may refuse to help out their neighbors who are in need. How does the community keep default or delinquency at bay? It seems that discipline is maintained by the pressure of community opinion and the threat of censure or ostracization—instruments that can be more powerful than mere pecuniary penalties. Note, moreover, that an impersonal market, by the very fact of being *impersonal*, does not have these instruments at its disposal. What an impersonal market does share with this informal institution is another source of punishment: temporary or permanent exclusion from future participation in reciprocal transactions.

In the few cases of dispute and perceived default, a complaint is made to a community leader or village head by the lender. If, after review of the situation, the leader holds the borrower to be guilty, the punishment imposed is usually a verbal admonition and, in extreme cases, a threat to make the matter public (leading to shame and possible sanctions from the community). The latter threat is considered so severe and damaging that in his survey, Udry found not a single instance of it.

Other issues of moral hazard are likely to be secondary in the Nigerian context because of the ease of information flow. Because lending is restricted within closely knit communities, borrowers usually find it impossible to misrepresent their true financial situation or to feign crop damage, illness, or any other misfortune.

A glance at default figures in the sample is revealing. Of the 808 loans, only 47 (about 5%) were in default as perceived by the lenders themselves. Astonishingly, only three of these cases pertained to borrowers who experienced negative shocks to their income after the loan was taken. This must mean that borrowers who do not repay because of bad shocks are generally not *perceived* as having defaulted. This goes to show that it is part of

the norm that the repayment burden is reduced or excused for those borrowers who run into financial trouble: proof again of the insurance aspect of these transactions.

Let's end with some observations on intervillage credit transactions. Note that a shock to income can be of two types: an *idiosyncratic* shock, which affects a single household (e.g., an illness in the family, flooding or insect damage to the family's plot, etc.), and an *aggregate* shock, which can affect an entire village or region (e.g., rainfall shortage, the flooding of a river, etc.). Because most borrowing and lending is restricted to small communities and villages, it seems that the rural credit market should be efficient in pooling risk and providing insurance against the first kind of shock, but not against the second. Udry estimated that in his sample, about half (58%) of the variation in farm yields was caused by the latter kind of shocks, so this limitation must be serious.

However, although there is practically no direct outside lending (the informational disadvantage of outsiders and the inability to charge fixed, nonstate contingent interest rates in violation of Islamic law are surely the reasons), flows of funds between villages are not entirely absent. In northern Nigeria, the Hausa tradition of long-distance trading has given rise to a network of merchants. With widespread outside connections and intimate personal knowledge of people in their own resident villages, these merchants act as pipelines for outside credit to enter the village, while still providing loans on terms that do not violate the accepted standards of the community. Intervillage insurance, though not as prevalent as intravillage arrangements, does not appear to be altogether absent.

20.9. Summary

This chapter studied *insurance*: the provision of funds to smooth out unexpected variations in income or in consumption needs. We began with an example to show how insurance can be profitable, and used this example to distinguish between *self-insurance*, which is the smoothing of consumption using one's own assets, and *mutual insurance*, which requires interaction across a number of individuals. We observed that insurance possibilities are maximal when the fortunes of individuals are *negatively correlated* (I am doing well, on average, when you are not, and vice versa). Although negative correlation may be difficult to come by in practice, a weaker condition—the *independence* of incomes—also yields considerable scope for insurance. Insurance possibilities do diminish, however, as the fortunes of individuals become more and more *positively correlated* and vanish altogether if incomes and needs move entirely in tandem across individuals. Finally, we observed that the larger the group, the greater the possibilities of mutual insurance.

We then introduced a benchmark model, which we call the *perfect insurance* scenario. In this scenario there are a large number of people who face an aggregate shock that is common to all of them (say, the weather) and the additional possibility of idiosyncratic variations in income (or consumption needs). We showed that it is optimal to pool all the idiosyncratic variation and that it will be impossible to insure against any of the aggregate variation. Hence, the solution to the perfect insurance problem yields a testable hypothesis: if the model is good, then we should see that individual (or household-level) consumption is unaffected by *individual* incomes, but it should move in perfect correlation with the aggregate consumption of the insurance group (the entire village, let us say).

We turned to regression analysis to test this implication. Use household consumption as the dependent variable and "group consumption," household income,

and other household observables (such as a spell of unemployment) as independent variables. Roughly speaking, if the theory of perfect insurance is right, the estimated coefficients on all the household-specific variables should be zero and the coefficient on group average consumption should be unity. We noted several problems with this approach that largely stemmed from missing data: for instance, what do we use for group consumption when all we have are data on a sample of households? We then discussed several empirical studies. The consensus is that although the coefficient on household income is generally small (certainly way smaller than unity), we are pretty far from perfect insurance. Idiosyncratic shocks that affect household income do have a significant effect on the consumption of that household.

This conclusion motivated a study of models with *imperfect* insurance. We identified two broad factors that might impose constraints on insurance: *limited information* and *limited enforcement*.

The informational problem is one of *moral hazard*: group members may not be sure that a particular member has suffered an income loss or a consumption spike; that is, it might be difficult or impossible to verify the outcome, on which insurance transfers are predicated. At a more subtle level, it may be possible to verify the outcome, but not what led to it: did a farmer suffer a bad harvest because he was really or truly unlucky or did he apply inadequate levels of effort to cultivation? With these uncertainties, it might be foolhardy to attempt perfect insurance: there would be huge efficiency losses in production because individuals would then have the incentive to apply inadequate effort to income-generating activities. We studied the imperfect insurance problem under these constraints and described a "second-best" insurance arrangement, suitably constrained by the limited information.

One implication of this model is that individuals run into a trade-off between diversification and moral hazard. Diversified, uncorrelated incomes across group members are easiest to achieve when members are all in different occupations or they are in the same occupation, perhaps, but are geographically separated, which is hard to accomplish within a village. On the other hand, an extension of insurance schemes across villages runs straight into the moral hazard problem, because it is much harder to verify outcomes and inputs. We discussed how this trade-off can be ameliorated in practice.

The second problem has to do with enforcement. A farmer with a good harvest may be required to make a transfer to the common insurance fund. Why would he do so? One answer is that if he does not comply, he will face social sanctions from the village. He might also be excluded from future participation in the insurance scheme. Such sanctions and exclusions have costs, but they must be traded off against the benefits from retaining extra income today. This notion gives rise to another sort of constrained problem: find the best insurance scheme consistent with the requirement that all farmers will pay up when times are good. We analyzed the properties of this scheme.

There are several implications of the enforcement-constrained model. One is that increased opportunities for self-insurance tend to destroy schemes of mutual insurance, because such opportunities lower the punishment value of future exclusion. Individuals are more likely to deviate from an ongoing scheme. Likewise, the loss of social ties

in the village weakens the power of social disapproval: this, too, will cause insurance schemes to fall apart.

Another implication of the enforcement-constrained model is that it generally predicts the blurring of credit and insurance. We observed that pure insurance is "history-free": the fact that I make a transfer today has no implications for what might happen tomorrow. However, when pure insurance is ruled out by problems of enforcement, a second-best compromise might be to make my future returns from the scheme positively related to transfers that I make today. The reason is that this eases the no-deviation constraint and allows for better insurance overall. The linking of future returns to current payouts can be intrepreted as a form of credit.

We ended the chapter with a discussion of rural markets in Nigeria, where this sort of mixed scheme can be seen in practice.

20.10. Summary

We studied the main features of informal credit markets. Typically credit is taken to finance ongoing production (*working capital*) or to finance shortfalls in consumption (*consumption credit*). Credit may be sought for a new venture, such as the overhead expenditures needed to switch to a new crop (*fixed capital*).

Two fundamental problems characterize credit markets: the problem of "inability to repay" (*involuntary default*) and the problem of "unwillingness to repay" (*strategic default*). In turn, both these problems are outcomes of informational imperfections. For instance, if the use of a loan cannot be monitored, it may be frittered away in consumption or in an overly risky project, leading to involuntary default. The fear that this might happen could shut down a potentially profitable credit transaction. Likewise, the absence of a legal system to honor credit transactions and a possible absence of information flow regarding defaulting borrowers might exacerbate the problem of strategic default.

We distinguished between institutional or *formal lenders*—government or commercial banks, credit bureaus, and so on—and *informal lenders*—village moneylenders, landlords, traders, and shopkeepers. Two considerations typically put formal lenders at a disadvantage relative to their informal counterparts: (1) they lack information regarding the characteristics, history, or current activities of their clients, and (2) they cannot accept collateral in some nonmonetary forms, such as labor or output. One example of the informational problem is a *limited liability* situation, in which a borrower repays if successful, but is otherwise limited by her resources to fully repay the loan. In such situations, a loan that cannot be properly monitored may be used in overly risky activities, because the borrower does not internalize the downside risk, which is borne by the bank (due to the limited liability).

We then turned to informal markets. Some characteristics of these markets are (1) *informational constraints* (which create repayment problems), (2) *segmentation* (in which lenders seem to deal with demarcated zones of borrowers), (3) *interlinkage* (in which lenders and borrowers are simultaneously transacting with each other on at least one other market, such as land or output), (4) *interest-rate variation* across borrower types, loan purpose, or spatial dimensions, (5) *credit rationing* (a borrower may wish

to borrow more at the going rate of interest in the transaction, but is not permitted to), and (6) *exclusivity* (a borrower typically deals with one lender at a time).

We discussed theories that throw light on some of these characteristics. Among theories of high interest rates are those that use the notion of *lender's risk*, which create risk premiums, and theories that view credit as a veil for transferring collateral (such as land) from borrower to lender. Particular attention was placed on issues of default: in particular, we argued that the possibility of default limits the granting of fixed-capital loans that lowers the borrower's dependence on future credit, and that strategic default may create credit rationing. Credit rationing may also be the outcome of informational asymmetries: interest rates cannot be made too high because of the fear of attracting only risky borrowers who do not internalize downside risk. We studied in detail the relationship between default and information flow: societies with high mobility may lose information, thereby making it easier for a defaulting borrower to get loans elsewhere and thus shrinking the informal credit market as lenders become more wary of offering loans.

We then turned to *interlinkages*: the tying of a credit transaction to transactions in some other market, such as a land tenancy or the buying and selling of grain. Interlinkages may be a way to charge interest in hidden forms because the charging of explicit interest is banned by law or custom. Interlinked contracts may be a way to economize on the lender's cost of acquiring information about his borrower or on the cost of tracking or monitoring a loan. An interlinked contract may serve to reduce strategic default (for instance, a landlord may be well placed to lend to his tenant because he has the additional threat of removing the tenancy contract in case a loan is defaulted upon). Finally, interlinkages may be a way to maintain the "correct" relative price across a bundle of transactions, thus minimizing distortions. For instance, a high interest rate may discourage loan taking for production, but a subsidized interest rate *and* a low buying price for output may maintain the loan taking incentive while allowing a trader–lender to acquire surplus via the reduced buying price.

Our last topic covered some policy issues. Faced with the difficulties of reaching small borrowers through formal-sector credit, a government might follow the policy of expanding credit to traders and landlords in the expectation that they will pass the funds on to small borrowers and be forced by increased competition to offer better terms. We evaluated this argument and found that there are some theoretical reasons to be suspicious of this quick conclusion. The costs of loan administration and monitoring may increase with more lenders (because borrowers can more easily default on existing loans). This might *raise* interest rates instead of lower them. Expanded formal-sector credit might make it easier for informal lenders to collude (because it raises the capacity of each of them to fight a price war in the case of a deviation from strategic collusion). Additionally, an expansion of funds to lenders who can screen good borrowers easily may "contaminate" the borrower pool for other lenders, leading to a credit shutdown on their part.

Another policy response is to try to put the hidden information of local communities to work in innovative ways. The leading example of small-scale lending or *microfinance* is the Grameen Bank of Bangladesh. The Bank lends to groups of people instead of individuals. The entire group is denied future credit if one member defaults. Under such circumstances, group formation will use local information (unknown to the bank,

but available to potential borrowers). Would-be members who are bad credit risks will be excluded from the group, leading spontaneously to better targeting. In addition, borrowers will have the incentive to urge each other to take on safer projects (because they now bear more of the downside risk owing to the threat of future exclusion). We discussed these and some potential drawbacks as well. Finally, we evaluated the financial viability and performance of the Grameen Bank.

Exercises

Return to the example in Section 19.7.3. Show that if pM is held fixed, then a lowering of p will still increase the chances that the safe lower rate will be charged.

Turn this into a problem: Why Not Just Adjust The Interest Rate? One reaction to all this might be: just ignore questions of risk, let the borrower do what he wants, and simply adjust the interest rate so as to compensate the lender (in expected terms). That adjustment is often called the *risk premium*. A bit more formally, if p is the probability that a loan will be repaid, a moneylender's expected profit is $p(1+r)L$, where L is loan size and r the rate he charges. If the moneylender can get a rate of return i elsewhere by safely investing his money, the interest rate r that will compensate him is given by

$$p(1+r)L - (1+i)L = 0,$$

or

$$r = \frac{1+i}{p} - 1. \tag{20.14}$$

For instance, if the safe return gets the lender 5%, and there is a 50–50 chance of default, then (20.14) implies that $r = 110\%$.

This seems like an obvious point but like many obvious points it is often wrong. The reason is that the interest rate premium itself affects borrower behavior, and it may spark off a higher chance of default. Nothing gives us the right to assume that p is exogenous. In fact, the higher is r, the higher might be p, a theme to which we return later. Interest rate adjustments can sometimes look like a dog chasing its own tail, and there may be no end-point to that chase.

Turn this into a problem. Bhaduri (1977). Suppose that a small farmer is in need of a loan of size L, perhaps to tide over a family emergency. He approaches the local large landowner, who is known to lend money for such purposes. The landowner asks him to pledge land as collateral for the loan. Our farmer has a plot of land adjacent to that of the large landowner and this is what he pledges.

Now let us keep track of some of the relevant variables by introducing some notation. Let i be the interest rate charged on the loan L and let V_S (S for "small") be the (monetary) value that the small farmer places on his land. Likewise, let V_B (B for "big") be the value that the big landowner attaches to the same plot of land. Because the plot is adjacent, V_B is not a negligible value: one can certainly imagine cases in which V_B exceeds V_S. Thus we have an example here of the first kind of collateral, one which is of value to both parties.

Next, let us place a monetary value (just as we did in Chapter 10) on the loss to the farmer from default, over and above the loss of his collateral. Such losses may include the fear of not receiving future loans or even the threat of physical retribution. Summarize the money value of this loss by F.

When the time comes to return the loan, we can conceive of two possibilities:

(1) The borrower may be in a state of involuntary default: he simply does not possess the wherewithal to repay the money. In that case, he certainly loses the land, which passes into the hands of the large landowner.

(2) The borrower may contemplate willful default and take his chances with the landless labor market or even with migration to the city. The total loss to the borrower in this case is $V_S + F$, whereas the gain is that he gets to keep the principal plus interest that he owed. Thus the borrower will prefer to return the loan if

$$L(1 + i) < V_S + F. \tag{20.15}$$

Consider, now, the lender's preferences. Does he prefer to see his loan returned or does he prefer to keep the collateral? He will prefer his money back if

$$L(1 + i) > V_B. \tag{20.16}$$

Combining (20.15) and (20.16), we may conclude that loan repayment is in the interest of *both* parties only if

$$V_B < V_S + F, \tag{20.17}$$

which is a way of saying that the lender's valuation must not exceed the borrower's valuation by too much. In the special case where $F = 0$, so that collateral is the *only* way to force loan repayment, (20.17) says that the lender's valuation of the collateral must be less than that of the borrower.

Turn this around. Suppose that (20.17) does *not* hold, so that $V_B > V_S + F$. In this case, it follows that whenever the borrower prefers to repay the loan, the lender actually wants him not to do so! The lender would actually like the credit transaction to be an excuse to acquire the collateral (cheap). Thus collateral that is of high value to *both* lender and borrower may (paradoxically) result in credit transactions with excessive rates of default.[334]

In the present model, how might high default rates be precipitated? One way is to drive up the rate of interest so that (20.15) fails. In this case, if the borrower does not default involuntarily, the circumstances will induce him to do so willfully. The interest rate on such credit transactions will be high indeed, but the principal objective of the transaction is *not* to earn interest!

This story may be of some relevance in explaining why land inequalities rise in poor societies. Land is passed from poor to rich in lieu of unpaid debt; this much is to be expected. What the story teaches us, however, is that the debt contract may be written so as to deliberately induce these transfers.

By the way, land is only one example. Bonded labor is another. A loan may be advanced for the express purpose of ensuring a supply of cheap labor from the borrower. In this case, labor power acts as the collateral that is transferred in the end.

Finally, note that this sort of analysis works much better for consumption loans than production loans. With consumption loans (taken, say, for an illness in the family), the amount is often fixed and cannot vary with the rate of interest. With production loans, a high interest rate may be self-defeating because the borrower can scale back

[334] In a sense, these rates are illusory: they are provoked because the lender would like to drive the borrower into a state of default.

the amount of the loan by reducing the extent of his productive activity. This hints at a possible explanation of why interest rates may be high for some types of credit transactions but not for others.

(1) Explain the concept of a *risk premium* in words. This example will help you understand how a risk premium may be calculated. In each of cases (a)–(d), assume that you always have the option to keep extra money in the bank at a 10% rate of interest, with no fear of losing any of this money. For each case, calculate the minimum rate of interest and, therefore, the risk premium, at which you would lend $1,000 on the informal market.

(a) With probability 1/2 the loan will be repaid with interest, and with probability 1/2 the loan will not be repaid at all.

(b) With probability 1/2 the loan will be repaid with interest, and with probability 1/2 only the principal will be repaid.

(c) With probability 1/3, the loan will be repaid with interest, with probability 1/3 only the principal will be repaid, and with probability 1/3 the loan will not be repaid at all.

(d) Just as in case (a), except that if the loan is defaulted upon, there is probability 1/2 of recovering assets from the borrower worth $500.

(2) If people have different costs of default, such that for each person there is a threshold repayment burden that they will always want to default upon, explain why loans above a certain size will never be offered by any lender, even with very high risk premiums built into the rate of interest. Use graphs to explain your answer.

Here is a special example that builds on the preceding general question. The following table gives default risks for various loan sizes in an informal credit market.

Loan size ($)	*Loans defaulted (%)*
50–99	5
100–149	10
150–199	20
200–249	25
250–300	30
>300	502

Suppose that the rate of interest in the informal sector is 18% per year and that in the formal sector (with no default) is 10% per year. Calculate the maximum loan size that will be offered in the informal-sector credit market. For what minimum rate of interest will loans in the $250–300 category be offered? (This example is more special than the general question in the first paragraph, because the default rates may also depend on the rate of interest being charged, and this is not taken into account in the example. How would you try to amend the example so that it fits the more general scenario?)

(3) A *risk-lover* is an individual who prefers a lottery with some expected value to a fixed amount of money with the same value. Show that with limited liability, a risk-neutral or even a risk-averse borrower may behave *as if* he is a risk-lover when considering the projects in which he would like to invest. Explain how this feature (of

limited liability) drives a wedge between projects that the lender considers profitable and those that the borrower considers profitable.

(4) Carefully reproduce the example of a landlord lending to a laborer, where underpriced labor as well as interest can be charged as repayment for the loan. Show that the loan will be advanced at the opportunity cost of the lender, whereas all interest will be charged by underpricing labor.

(5) Carefully reproduce the example of a trader lending to a farmer, where output sold by the farmer to the trader at a discounted price (as well as direct interest) can be required as repayment for the loan. Show that the loan will be advanced at a rate of interest that is *below* the opportunity cost of funds for the trader, and all profits are made by the trader by buying output at a discounted price.

(6) Here is another way to think about output–credit interlinkage between a trader and a farmer. Recall that the borrower–farmer maximizes *his* profit. Note that (i) marginal revenue equals price times marginal product of an extra \$1 of loan and that (b) marginal cost equals \$1 plus the rate of interest on \$1.

(a) Show that if MP is marginal product, p = price, and i is the rate of interest, then $\text{MP} = (1+i)/p$.

(b) We argued that the optimal contractual terms must not distort the size of the loan. Show that if p^* is the price contracted by the trader and i^* is the rate of interest charged by the trader,

$$\frac{1+i}{p} = \frac{1+i^*}{p^*}.$$

(7) Do you expect that there will be a systematic discrepancy in interest rates between loans that are made for production purposes and loans that are given for consumption purposes? Analyze some sources of this possible discrepancy.

(8) An expansion of credit from formal financial institutions to large informal lenders will increase competition among such lenders and improve the terms of credit for small borrowers with no access to formal credit. Discuss.

(9) Consider a monopolist lender who lends to borrowers on a repeated basis. The loans are informal and are not backed up by written contracts. The lender has no way to recover a loan if the borrower chooses to default. The lender, however, threatens to cut off credit in the future to any defaulting borrower. Borrowers discount the next period's earnings by a discount factor of 0.5 (see Appendix 1 on game theory at the end of the book for a discussion of this concept).

Borrowers use the loan in cultivation. Cultivation can be done using one of two techniques. The first requires initial working capital of \$100 and produces net output worth \$300. The second requires \$500 of working capital and yields net output of \$1,000. Find the amount of loan the lender will advance to each borrower every period in order to maximize his own profits. How much is the scheduled repayment and the implicit interest rate? What are these profits? How much does the borrower earn every period from the deal?

Introduce a new factor into this scenario. Suppose the lender can keep some of the borrower's assets (like jewelry) as collateral, which he will seize in the case of default. The present value of the asset to the borrower is \$300. Recalculate the optimal loan, repayment amount, implicit interest, and profits in this case. Compare the two cases

and summarize the effect of collateralization on the other terms of the loan. Does it increase or decrease the welfare of the borrower and the lender?

(10) SelfHelp is a newly formed credit cooperative which receives partial financing from government banks. Members of SelfHelp can deposit savings with the cooperative and they can also turn to SelfHelp for a loan if they need one. If a borrower defaults on a SelfHelp loan he is punished (which is equivalent to a loss of monetary value F) and excluded from future dealings with SelfHelp. The value (to the borrower) of these future dealings is some number S. However, there is no telling whether SelfHelp will survive in the future; denote the probability of survival by p.

(a) If each member is risk-neutral, what is the *expected* value of dealing with SelfHelp in the future?

(b) If a member has an outstanding loan of L and needs to repay it (along with interest at the rate r), write down the value of his *net* gain from default.

(c) Now suppose that the probability of SelfHelp's survival depends on the percentage of borrowers who repay as well as the quantum of government assistance should a high rate of default occur. Draw a graph with the repayment rate on one axis and the survival probability on the other, and show how this graph shifts with the amount of government assistance.

(d) Use parts (b) and (c) to show that in general there can be two outcomes or equilibria for the *same* parameters: (i) there is no (voluntary or strategic) default and SelfHelp survives with little or no government assistance or (ii) there are high rates of default and SelfHelp survives with low probability despite government assistance.

(e) Show that a credible promise by the government to bail out SelfHelp in times of trouble can lead to a unique equilibrium in which all borrowers repay and little government assistance is *actually* required [scenario (i) in part (d)].

(f) Using part (b), show how the survival probability of SelfHelp can affect the size of the loans it can make to borrowers.

(11) A lending organization, inspired by the Grameen Bank, is attempting to provide loans to small farmers. It is lending to farmers in groups of two (say).

(a) Provide at least two reasons why a strategy of group lending may be better than a strategy of lending to individuals. Provide at least two reasons why it may be worse.

(b) A sequential lending strategy is one in which group members are given loans in some order, with the next member receiving a loan only after the earlier member has repaid. A simultaneous lending strategy is one in which all group members are given loans at the same time. In both cases, assume that default by any one member blacklists the whole group. Compare and contrast these two strategies.