

Inequality and investment

chapter 5

In a world in which markets worked perfectly, investment decisions would have little to do with the income, wealth, or social status of the decision maker. They would be determined by the returns an investment promises and by the market price of capital, adjusted for the extra risk it entails. If people had good investment opportunities, it really would not matter whether they had the money—they could always borrow what they needed, and if the risk bothered them, they could always sell shares in their business and buy safer assets with the money from the sale.

However, for various reasons—mainly economic but also political—markets are not perfect. If borrowers can willfully default on their loans, lenders prefer to make loans to borrowers who can provide collateral assets. Private returns for politically connected firms can be higher than for those without such connections, and so these firms may attract more capital, even though social returns may not be any greater.¹ Members of groups subject to discrimination may rationally invest less in their human capital than they would in the absence of such explicit or subtle stereotypes.

After we give up the idea that markets work anywhere close to perfectly, the scope for a direct link between investment and the distribution of wealth or power widens substantially, in many instances leading to underinvestment by those who have good growth opportunities.² Correcting the market failures directly is often not feasible, and in these cases certain redistributions of wealth, power, and resources can serve as second-best alternatives.³ In other words, interventions to enhance equity can improve efficiency.

One of the great advances in development economics in the past 15 years is the accretion of a substantial body of evidence on documenting how well (or badly) asset and financial markets work in developing countries. The fact that these markets rarely measure up to their ideal creates the possibility that wealth and social status, defined as one's position in society both in ascriptive identity and in connections, will have an important influence on investment decisions. It seems natural to start with this evidence.

Markets, wealth, status, and investment behavior

The market for credit

In a perfect credit market, there is a single interest rate and everyone can borrow or lend as much as they want at that rate. That individuals can borrow as much as they want at the current rate explains the presumption of a separation between the wealth or status of the investors and the amount they invest. Whether they are rich or poor, well-connected or just off the streets, an extra dollar of investment will be profitable for them only if the return they get from it is more than the interest rate. If the interest rate is higher, they would be better off lending that money if it was their own, or borrowing less if it were someone else's. So, two people with the same return on investment would end up investing the same amount.⁴

How close are real markets to this ideal market? Chambhar is a market town in Sindh (Pakistan), on the east bank of the Indus. In 1980–81 farmers from the area around Chambhar got most of their credit from about 60 professional moneylenders. Based on detailed data from 14 of these

lenders and 60 of their clients, Aleem (1990) calculated the average borrowing interest rate charged as 78.5 percent. But if these farmers wanted to lend their money, the banking system would pay them only about 10 percent. It is possible, however, that they may not have been depositing in the banks. An alternative measure of the deposit rate that is relevant for these farmers is the opportunity cost of capital to the moneylenders, 32.5 percent. In either case, it suggests a gap of at least 45 percentage points between the borrowing and lending rates.

The borrowing rate also varied enormously across borrowers. The standard deviation of the interest rate was 38.1 percent, compared with an average lending rate of 78.5 percent. In other words, an interest rate of 2 percent and an interest rate of 150 percent are both within two standard deviations of the mean. One possibility is that these differences in interest rates reflect differences in the default rate: perhaps the expected repayment was the same for everybody, because those who paid higher rates were more likely to default. Also the expected repayment could have been equal to the actual interest rate paid to the depositors, if the default rate was high enough. But default was rare: for individual lenders, the median default rate was between 1.5 percent and 2 percent, with a maximum of 10 percent.

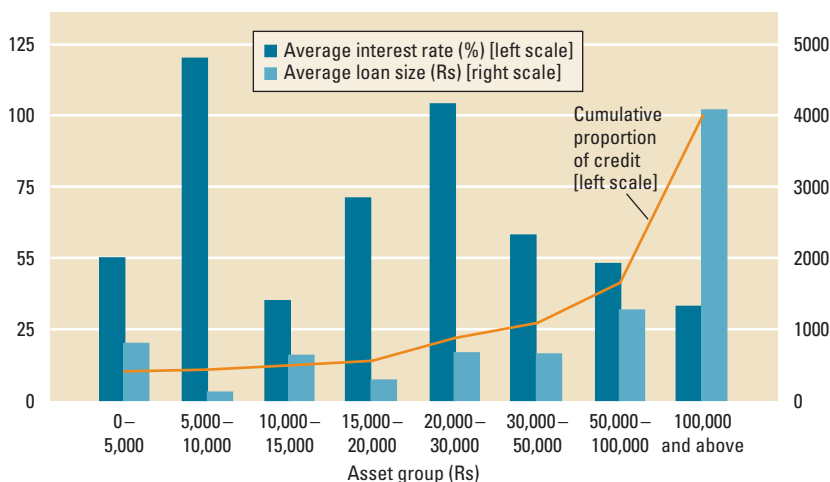
The same pattern—high and variable borrowing rates, much lower deposit rates, and low default rates—shows up in the “Summary Report on Informal Credit Markets in India.”⁵⁵ This report summarizes results from case studies commissioned by the Asian Development Bank and carried out under the National Institute of Public Finance and Policy.

For the urban sector, the data are based on various case surveys of specific classes of informal lenders. For the broad class of nonbank financial intermediaries called finance corporations, the maximum deposit rate for loans of less than one year is 12 percent. These corporations offer advances for one year or less at rates that vary from 48 percent per year to the utterly astronomical 5 percent per day. The rates on loans of more than one year varied between 24 percent and 48 percent. Default, once again, is only a small part of the story: default costs explain only 4 percent of total interest costs. For hire-purchase companies in Delhi, the deposit rate was 14 percent and the lending rate was at least 28 percent and could be as high as 41 percent. Default costs were 3 percent of total interest costs.

For the rural sector, interest rates are high, but they are also variable (figure 5.1). This finding is based on surveys of six villages in Kerala and Tamil Nadu, carried out by the Centre for Development Studies, Trivandrum. The rich (with Rs 100,000 or more in assets) get most of the credit (nearly 60 percent) and pay a relatively low rate (33 percent), while those with assets between Rs 20,000 and Rs 30,000 pay rates of 104 percent and get only 8 percent of the credit. The average interest rate charged by professional moneylenders (who provide 45.6 percent of the credit) is about 52 percent.

While the average deposit rate is not reported, the maximum from all the case studies is 24 percent, and in four of them it is no more than 14 percent. In the category of professional moneylenders, about half the loans were at 60 percent or more, but another 40 percent or so had rates below 36 percent. Default rates were higher than in the urban sector, but they still cannot explain more than 23 percent of the interest costs.

Figure 5.1 In rural Kerala and Tamil Nadu, the rich access most of the credit and pay relatively low rates



Source: Dasgupta, Nayar, and Associates (1989)

The fact that credit access depends on social status is also shown by Fafchamps' (2000) study of informal trade credit in Kenya and Zimbabwe. It reports an average monthly interest rate of just over 2.5 percent (corresponding to an annualized rate of 34 percent), but it also notes that the rate for the dominant trading group (Indians in Kenya, whites in Zimbabwe) is 2.5 percent a month, while the blacks pay 5 percent a month in both countries.⁶ Chapter 9 also provides evidence that in many countries "insiders" effectively lobby to limit access to financial institutions and that lending is skewed toward the rich, consistent with the evidence in figure 5.1.

None of these facts is surprising. Contract enforcement in developing countries is often difficult, and it is not easy to get courts to punish recalcitrant borrowers.⁷ As a result, lenders often spend a lot to make sure that their loans get repaid: it is plausible that these are the resources that drive a wedge between the borrowing rate and the lending rate. Indeed, Aleem (1990) shows that the resources spent by lenders to monitor borrowers explain the nearly 50 percentage point gap between the lending and borrowing rates in his data. It is easy to imagine that borrowers who are easier to monitor will enjoy better rates, which would explain why lending rates vary so much.

These imperfections in credit markets have immediate implications for the relationship between wealth and investment. First, with the rate of interest on deposits much lower than that on loans, the opportunity cost of capital for those who just want to invest their own money is much lower than the opportunity cost for those who have to borrow. This means that the wealthy will end up investing much more than the indigent, even if they face exactly the same returns on their investment. Second, the lower interest rates charged to rich people reinforce this conclusion, because the rich then face a lower opportunity cost when they too are borrowing. Third, in some cases, those who are unable to provide collateral will have no access to credit at any interest rate.

We would thus expect the poor to underinvest, certainly relative to the rich, but also

relative to what would happen if markets functioned properly. The capital released because they underinvest is absorbed by the non-poor, who may actually end up overinvesting relative to how they would invest in perfect markets. This is the reason: because the poor cannot borrow, the non-poor cannot lend as much as they would like to (this is why deposit rates in developing countries are often very low). And because the non-poor cannot lend, it makes sense for them to keep investing in their own firms, even when the returns are low.

Because the poor underinvest, and because the opportunity cost of capital to the non-poor is thus lower than it would otherwise be, the composition of the investors also changes. In particular, firms that would not be viable if markets functioned perfectly (for example, because the interest rate would be too high) can survive and even expand because markets are the way they are. In other words, the "wrong" firms end up investing.

The market for insurance

The ideal insurance market is one in which people bear no avoidable risks. In a setting in which a single village constitutes a separate insurance market closed to the rest of the world (so that only people in the village can insure other people in the village, in some kind of mutual insurance arrangement), individual consumption should respond only to aggregate (village-level) income fluctuations and not to fluctuations in the income of specific individuals. Put in blunter terms, as long as aggregate consumption is unchanged, individual income fluctuations should not translate into fluctuations in individual consumption. When insurance markets work well, risk considerations should not have a significant impact on the choices people make, irrespective of their wealth, given that what an individual does has little impact on aggregate uncertainty.

While a perfect insurance market is more complex than a perfect credit market, and thus harder to detect, there have been attempts to test the prediction about the irrelevance of fluctuations in one's own income. The Côte d'Ivoire Living Standards Measurement Surveys from 1985 to 1987

provide panel data on the income and consumption of nearly 800 households, with each household tracked for two consecutive years (1985 and 1986 or 1986 and 1987). In table 5.1, the relationship between changes in consumption and changes in incomes is reported separately for the three main regions and separately for 1985–86 and 1986–87. The first row of the first block for each year reports the basic correlation between income and consumption: a fall in income always hurts consumption, although the coefficient varies between a low of 0.15 (a \$1 reduction in income means that consumption goes down by \$0.15) to a high of 0.46. The next row does the same thing, but now there is a village dummy intended to pick up any village-level changes in consumption. Remarkably, the coefficients on own income, which under perfect insurance should have fallen to zero after controlling for village-level changes, barely budge.⁸

Not all the evidence is quite so pessimistic. Townsend (1994) used detailed household-level data from four villages, which were intensively studied by the International Crop Research Institute in the Semi-Arid Tropics (ICRISAT) in India, to see whether the full insurance hypothesis is consistent with the data. He found that while the data did reject the exact prediction, it did not miss by very much. In other words, his evidence suggested that villagers do insure each other to a considerable extent: movements in individual consumption in his data seem largely uncorrelated with movements in income.

Later work by Townsend, based on data he collected in Thailand, turned out to be less encouraging.⁹ Some villages seemed to

be much more effective than others in providing insurance to their residents. Townsend describes in detail how insurance arrangements differ across villages. While in one village there is a web of well-functioning, risk-sharing institutions, the situations in other villages are different. In one village, the institutions exist but are dysfunctional; in another, they are nonexistent; in a third, close to the roads, there seems to be no risk-sharing whatsoever, even within families.¹⁰

As for credit, the failure of insurance could have something to do with informational asymmetries. It is not easy to insure someone against a shock that he alone observes, because he has every incentive to always claim that things had gone badly. But as Duflo and Udry (2004) demonstrate, spouses in Côte d'Ivoire do not seem to be willing to insure each other fully against rainfall shocks that affect them differentially. Because rainfall obviously is observable, at least part of the problem has to be elsewhere. One possibility is limited commitment. People may be happy to claim what was promised to them when it is their turn to be paid, and then default when the time comes for them to pay. This may be particularly easy in a setting in which the social relations between the sets of people who are insuring each other are not particularly close, perhaps explaining why Townsend found no insurance in the village closest to the road.

Lack of insurance should have an effect on the pattern of investment. That many insurable risks are uninsured means that one cannot invest without personally bearing a significant part of the concomitant risk. Indeed, big corporations able to sell

Table 5.1 The effect of income shocks on consumption, Côte d'Ivoire

	West Forest		East Forest		Savannah		All Rural	
OLS 1985–6								
No dummies	0.290	(6.2)	0.153	(3.2)	0.368	(5.8)	0.259	(8.8)
Village dummies	0.265	(5.7)	0.155	(3.5)	0.373	(5.7)	0.223	(7.7)
OLS 1986–7								
No dummies	0.458	(8.8)	0.162	(5.3)	0.168	(4.0)	0.239	(10.4)
Village dummies	0.424	(8.1)	0.173	(5.6)	0.164	(3.8)	0.235	(10.1)

Source: Adapted from Deaton (1997), table 6.5, 381.

Note: Absolute value of t-statistics are shown in brackets. The first row of each panel shows the coefficient on income change of a regression of consumption changes on income changes. The second row reports the same result when village dummies are included in the regression. OLS = Ordinary Least Squares.

their equity in organized equity markets may be the only players who can really hope to diversify away a large part of the risk of a particular project. Given this fact and the reasonable assumption that the poor are more risk-averse than the rich, we are likely to be in a perverse situation in which the poor may also find it hardest to reduce their exposure to risk. Thus, they are likely to shy away from riskier and higher-return investments, reinforcing the prediction that the poor invest too little.

The market for land

In a perfect land market, individuals can buy or lease as much land as they want for as long as they want at a price that depends only on the quality of the land (and the length of the lease). The lease should be at a fixed rent, so that the lessor is the residual claimant on the produce of the land. That land can be freely bought and sold ensures that there is no particular advantage or disadvantage to owning land compared with any other asset of similar value. That the lessor is a residual claimant means that the land is put to optimal use. Not so, however, in practice.

Many developing (and some developed) countries have regulations about who can buy land and how much or how little they can buy. Binswanger, Deininger, and Feder (1995) argue that almost every developing country today has gone through a phase when it had regulations intended to concentrate landownership. By contrast, Besley and Burgess (2000) provide a list of regulations from different states in India, each of which is an attempt to limit the concentration of ownership in land.

Governments also directly limit transactions in land, with the ostensible aim of preventing the accumulation of land in the hands of a few people. In Ethiopia in the late 1990s, Deininger and others (2003) note that selling and mortgaging land were against the law. While rentals were officially allowed (after being disallowed for two decades), local leaders and governments were free to restrict even these rental transactions in land. For example, the Oromia region allowed farmers to rent only 50 percent of their holding and stipu-

lates maximum contract terms of 3 years for traditional technologies and 15 years for modern technologies.

It is often unclear who has the right to sell a particular plot of land, when no single person or family has a clear, undisputed, legal title to the land. This ambiguity reflects encroachments and land grabs in the evolution of land rights, as well as the importance of custom in governing land relations, especially in Africa. The recent popularity of land titling as a social intervention is a direct consequence.

Where lease contracts exist, they are not always of the fixed-rent type, at least when the land is used for cultivation. Many countries, including the United States, have a long tradition of an alternative contractual form: sharecropping. Under sharecropping, the farmer gets only a fraction of the produce, but he does not need to pay a fixed rent. As Alfred Marshall pointed out more than one hundred years ago, this weakens incentives and reduces the productivity of the land, but the near universality of sharecropping suggests that it is a response to a real need. There is some disagreement among economists about the exact nature of that need.¹¹ It is plausible, however, that the need is related to the fact that farmers are often poor, and making them pay the full rent when their crop does poorly is difficult and probably not desirable.

Leaseholds in developing countries tend to be short-lived. The norm is either a year or a season. Longer leases are not unknown, but they are rare. This might reflect the fact that custom, rather than law, secures most of these leases; perhaps it is too much to rely on custom to enforce leases of arbitrary length.

The imperfect salability of land can, of course, hurt anyone who owns it. But the rural poor probably have more of their wealth in land than most people, so making land nonsalable might be particularly harsh on them.

What tends to discourage investment in the land is the lack of an explicit title, or the insecurity of tenure more generally (caused, for example, by the short duration of leases and the possibility that the landlord might threaten to take the land away at the end of the lease). It clearly helps if land is owned

by the person contemplating the investment. That most who work in agriculture tend to be too poor to buy out the land they are cultivating is thus a potential source of underinvestment.

The market for human capital

One thing makes the market for human capital different from all the other asset markets: many decisions about investing in human capital are made by parents (or other family members) for their children. In other words, those making the decisions are different from those who receive the human capital. It is not hard to imagine why this separation might introduce important distortions to the functioning of this market. Gary Becker's classic formulation avoids this issue by assuming that the family can borrow against the child's future income, turning the problem into a conventional investment decision. Under that assumption, the amount invested will not depend on the family's means.

In practice, however, although human capital is an asset, it cannot be legally pledged or mortgaged, for the simple reason that pledging your human capital would be tantamount to selling yourself into slavery.¹² This obviously constrains people's ability to borrow money to finance investments in their education.

When parents cannot borrow against their children's future income—true most of the time in most developing countries—they may still hope that those children will take care of them in their old age. The hope might be that the children do grow up to reap the benefits of their parent's investment and that they will pay their parents back. But children know that they have no legal obligation to do so. If they do repay their parents, it is because they love their parents or because society expects them to do so.

Investments in human capital may thus be driven as much by parents' sense of what is the right thing to do, as by any calculation of costs and benefits. Once we accept this, it becomes clear that children's human capital may not be very different from any other consumption good—so richer families will tend to invest more in their children's health and education. And human capital decisions

may be more a product of culture and tradition than of the cold calculation of benefits. Benefits are relevant, but the responsiveness to them may not be as large as one might have expected.

In the market for human capital, the reward should be based entirely on the human capital supplied, not on other attributes of the person supplying the skills. Discrimination based on gender, caste, religion, or race obviously violates this, but so does a system of job allocation based on contacts. Until very recently, job discrimination based on gender was the norm all over the world, and the number of countries where such discrimination is still either legally or socially accepted is dwindling but significant. Even where such discrimination is explicitly frowned on, there is some evidence of continuing discrimination. The same is true of race, caste, and religion. Most discrimination—unless legally mandated through affirmative action in favor of a historically disadvantaged group, such as low castes in India and African Americans in the United States—flies in the face of explicit laws against it.

One reason discrimination is so hard to eliminate comes from its sheer insidiousness. Beliefs about differences are embedded in everyday attitudes and practices in a way that neither the discriminator nor the discriminated against may be conscious of, even though these beliefs transform how they both behave. This is what underlies the power of the stereotype. In a telling example, Stone, Perry, and Darley (1997) asked all participants in a recent experiment (American Caucasians, hereafter referred to as whites) to listen to the same running account of an athlete's basketball performance on the radio. Half the participants were led to believe that the target player was white, half that he was African American. The results indicated that information was less likely to be absorbed if it was discordant with the prevailing U.S. stereotypes that whites are more academically talented than African Americans, and that African Americans are more athletically gifted. The white target player was perceived as exhibiting less natural athletic ability but more "court smarts." The African-American target player

was perceived as exhibiting fewer court smarts but more natural athletic ability.

Such biases have also been documented in real-world settings. A recent study of the effect of stereotyping on judgment finds that prison inmates with more Afrocentric features receive harsher sentences than those with less Afrocentric features, controlling for race and criminal history.¹³

Bertrand and Mullanaithan (2003) show evidence from a field experiment proving beyond reasonable doubt that there is a high degree of African-American discrimination in the United States. They sent the same resumes to a large number of companies under either a stereotypically white name or a stereotypical African-American name, and found a 50 percent higher callback rate when the name was white. The data say that having a white name is worth as much as eight additional years of job experience. Moreover, the discrimination tended to be greater when the resume corresponded to someone who was better educated, suggesting that investment in human capital among African Americans probably is significantly underrewarded.

A very different form of discrimination comes from the allocation of jobs based on contacts. Munshi (2003) presents persuasive evidence that contacts are very important in the allocation of jobs for migrant labor in the United States. The employment prospects for Mexican migrants there, it turns out, are much better when they are from areas where there was an earlier outflow of migrants. Quite remarkably, it helps if migrants are from an area where there was a drought several years ago, which pushed out a cohort of migrants to the United States. These migrants then help the later generations of migrants from that area to find jobs. This is the clincher: it does not help to be from an area where there was a recent drought.

The perception of discrimination, conscious or not, can affect investments in human capital. Those who expect to be discriminated against in a particular labor market—rightly or wrongly, consciously or otherwise—will tend to invest less in acquiring the type of human capital that the market rewards. This could, perversely, gen-

erate self-reinforcing behavior. If members of the discriminated group invest less in their own education, or in searching for employment, others might use this underinvestment to confirm their prejudice against that group.

Stereotypes can be self-fulfilling not only because they influence perceptions of the target of the stereotype, but also because they influence the behavior of the individuals who are stereotyped. Stone and others (1999) asked college undergraduate volunteers to play a miniature golf course. Performance was measured by how many strokes were needed to put the ball in the hole: fewer strokes meant better performance. The variable that the experimenters manipulated was the description of the task. In one treatment, the task was described as a “standardized test of natural athletic ability,” in the other as a “standardized test of sports intelligence.” When the task was described as a test of natural athletic ability, the African-American participants performed better than the whites: they averaged 23.1 strokes to complete the 10-hole golf course, compared with 27.8 for the whites. But when the task was described as a test of sports intelligence, the race gap was reversed: African Americans averaged 27.2 strokes, whites 23.3.

One way to interpret this behavior is that social ideas—stereotypes about the talents of different social groups—impose bounds from within. Under the rational, self-interest hypothesis, individuals change their behavior only when their preferences or external constraints change. But the behavior of real individuals depends as well on belief systems that society impresses on them. Negative stereotypes create anxiety that may interfere with performance: that is why the psychologist Claude Steele termed this kind of behavior “stereotype threat.”¹⁴ The beliefs underlying the stereotypes, if deeply internalized, can affect early decisions about prospective careers, and attitudes toward society, by changing what Appadurai (2004) calls a person’s “capacity to aspire.” The reader may recall the example (from chapter 2) of the Batwa girl who wanted to be a cleaner upon completing school. Positive stereotypes, by contrast, can

boost self-confidence and lead individuals to expend greater effort.

Stereotypes influence behavior twice—through their impact on individuals' self-confidence, and through their impact on the way individuals expect to be treated. To examine the effect of stereotypes on the ability of individuals to respond to economic incentives, Hoff and Pandey (2004) undertook experiments with low- and high-caste children in rural north India. The caste system in India can be described as a highly stratified social hierarchy in which groups of individuals are invested with different social status and social meaning.

In the first experiment, groups composed of three low-caste ("untouchable") and three high-caste junior high school students were asked to solve mazes and were paid based on the number of mazes they solved. In one condition, no personal information about the participants was announced. In a second condition, caste was announced with each participant's name and village. In a third condition, participants were segregated by caste and then each participant's name, village, and caste were announced in the six-person group.

When caste was not announced, there was no caste gap in performance (figure 5.2). But increasing the salience of caste led to a significant decline in the average per-

formance of the low caste, regardless of whether the payment scheme was piece rate (that is, participants were paid 1 rupee per maze solved) or tournament (that is, the participant who solved the most mazes was paid 6 rupees per maze solved, while the other participants received nothing). When caste was announced, the low-caste children solved 25 percent fewer mazes on average in the piece-rate treatments, compared with the performance of subjects when caste was not announced. When caste was announced and groups were composed of six children drawn from only the low caste (a pattern of segregation that for the low caste implicitly evokes their traditional outcast status), the decline in low-caste performance was even greater. While we cannot be sure from these data what the children were thinking, some combination of loss of self-confidence and expectation of prejudicial treatment likely explains the result.

The expectation by the low-caste subjects of prejudicial treatment may be rational given the discrimination in their villages. But the discrimination itself may not be fully rational. Cognitive limitations may prevent others from judging stigmatized individuals fairly. That people are bounded in their ability to process information creates broad scope for belief systems—in which some social groups are viewed as innately inferior to others—to influence economic behavior. If such beliefs persist, it will generally be rational for those discriminated against to underinvest (with respect to others) in the accumulation of skills for which the return is likely to be lower for them. This rational calculation is additional to any reduction in their "capacity to aspire," arising from the internalization of those beliefs.

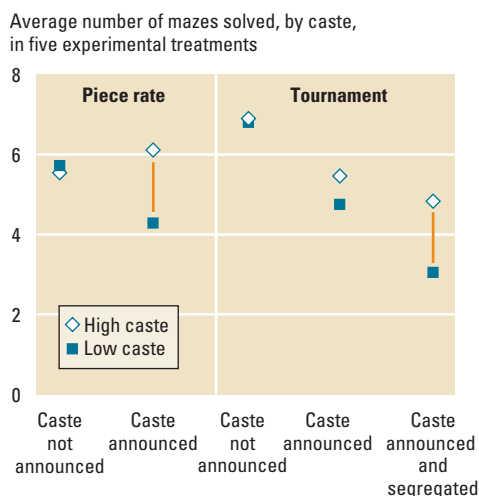
The evidence on underinvestment

Highly imperfect markets suggest considerable scope for underinvestment.

Industry and trade

Direct estimates of marginal products show that there are many unexploited investment opportunities. For small Mexican firms

Figure 5.2 Children's performance differs when their caste is made public



Source: Hoff and Pandey (2004).

Note: A vertical line in the figure indicates that the caste gaps are statistically significant.

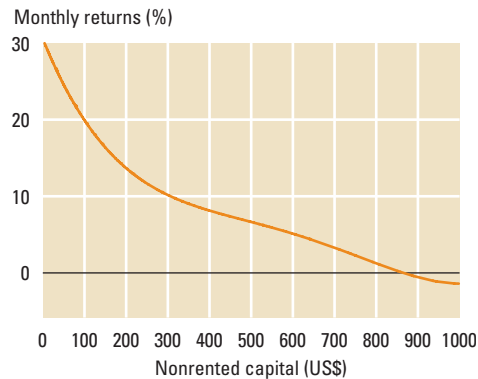
with less than \$200 invested, the rate of return reaches 15 percent per month, well above the informal interest rates available in pawn shops or through microcredit programs (on the order of 3 percent a month) (figure 5.3).¹⁵ Estimated rates of return decline with investment, but the rates remain high—7 percent to 10 percent a month for firms with investments between \$200 and \$500, and 5 percent for firms with investments between \$500 and \$1,000. All these firms are thus too small and could reap large gains from increased investment.

Trade credit is an important form of credit everywhere, perhaps especially where the formal institutions of the credit market are underdeveloped. Fisman (2001a) looked at the relation between access to trade credit and capacity utilization for 545 firms in Côte d'Ivoire, Kenya, Tanzania, Zambia, and Zimbabwe. He finds that firms that receive trade credit from three main suppliers (on average, about one of the three suppliers provides trade credit) have 10 percent better capacity utilization than firms that receive no trade credit. The relation is much stronger in industries in which it is important to carry large inventories.

Such studies present serious methodological issues, however. The basic problem comes from the fact that investment levels are likely to be correlated with omitted variables. For example, in a world without credit constraints, investment will be positively correlated with the expected returns to investment, generating a positive “ability bias.”¹⁶ McKenzie and Woodruff (2003) attempt to control for managerial ability by including the firm owner's wage in previous employment. This goes only part of the way, however, if individuals choose to enter self-employment precisely because their expected productivity in self-employment is much higher than their productivity in an employed job. Conversely, if capital is allocated to firms to avoid their failure, there could be a negative ability bias.

Banerjee and Duflo (2004a) take advantage of a change in the definition of the “priority sector” in India to circumvent these difficulties. All banks in India are required to lend at least 40 percent of their net credit to the “priority sector,” which

Figure 5.3 Returns to capital vary with firm size: evidence from small Mexican firms

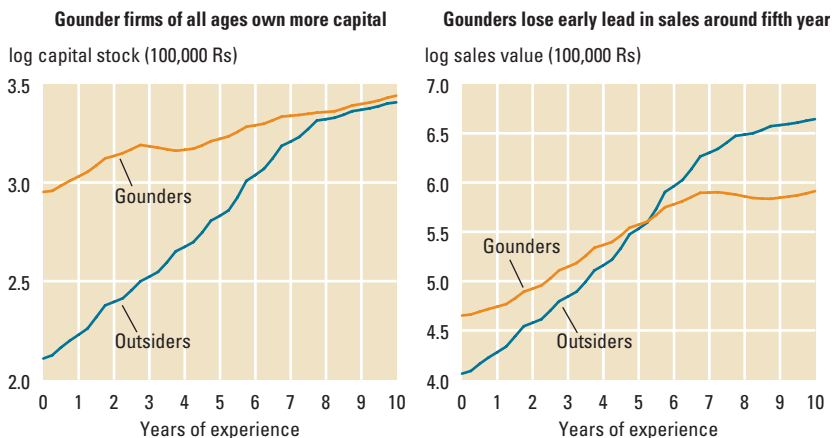


Source: McKenzie and Woodruff (2003).

includes small industry. In January 1998, the limit on total investment in plants and machinery for a firm to be eligible for inclusion in the small industry category was raised from Rs 6.5 million to Rs 30 million. The researchers first show that, after the reforms, newly eligible firms (those with investment between Rs 6.5 million and Rs 30 million) received, on average, larger increments in their working capital limit than smaller firms. They then show that the sales and profits increased faster for these firms during the same period. Putting these two facts together, researchers can estimate the impact of the increased access to working capital on the growth in profits. Allowing for the possibility that the firms in the priority sector were paying less than the true cost of capital for the extra money from the bank, they estimate that the returns to capital in these firms must be at least 94 percent.

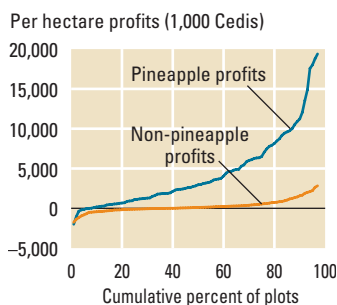
A different kind of evidence for underinvestment comes from the fact that many people pay the high interest rates reported earlier. Given that this money typically goes into financing trade and industry, the presumption is that the people borrowing at these rates of often 50 percent or more must have a marginal product of capital that is even higher. But the average marginal product in developing countries seems to be nowhere near 50 percent. One way to get at the average of the marginal products is to look at the incremental capital-output ratio (ICOR) for the country as a whole.¹⁷ For

Figure 5.4 Inefficient allocation of resources; the example of the Gounders vs. the outsiders



Source: Banerjee and Munshi (2004).

Figure 5.5 Average returns for switching to pineapples as an intercrop can exceed 1,200 percent



Source: Goldstein and Udry (1999).

the late 1990s the International Monetary Fund (IMF) estimates the ICOR to be more than 4.5 for India and 3.7 for Uganda. The implied upper bound on the average marginal product is 22 percent for India and 27 percent in Uganda.

That many firms in India have a marginal product of 50 percent or more, while the average marginal product is only 22 percent or so, is strong *prima facie* evidence for the misallocation of capital. The firms with the marginal product of 50 percent and more are clearly too small, while other firms (the ones who bring the average down to 22 percent) must, in some sense, be too large.

A specific example of this kind of misallocation of capital comes from a study of the knitted garment industry in the southern Indian town of Tirupur.¹⁸ Two groups of people operate in Tirupur: Gounders and outsiders. The Gounders, who issue from a small, wealthy, agricultural community from the area around Tirupur, moved into the readymade garment industry because there were not many investment opportunities in agriculture. Outsiders from various regions and communities started joining the city in the 1990s.

The Gounders, unsurprisingly, have much stronger ties in the local community, and thus better access to local finance. But they may be expected to have less natural ability for garment manufacturing than the outsiders, who came to Tirupur precisely because of its reputation as a center for garment export. The Gounders own about

twice as much capital as the outsiders on average. Gounder firms of all ages own more capital, although there is a strong tendency toward convergence as the firms age (figure 5.4a). The Gounders, despite owning more capital, lose their early lead in sales by about the fifth year, and end up selling less (figure 5.4b). In other words, outsiders invest less and produce more. They are clearly more able than the Gounders,¹⁹ but because they are less cash-rich and do not have the right connections, they end up working with less capital.

Agriculture

There is also direct evidence of high rates of returns on productive investment in agriculture. In the forest-savannah in Southern Ghana, cocoa cultivation, receding for many years because of the swollen shoot disease, has been replaced by a cassava-maize intercrop. Recently, pineapple cultivation for export to Europe offered a new opportunity for farmers in this area. In 1997 and 1998 more than 200 households cultivating 1,070 plots in four clusters in this area were surveyed every six weeks for about two years. Pineapple production dominates the traditional intercrop (figure 5.5),²⁰ and the average returns associated with switching from the traditional maize and cassava intercrops to pineapple is estimated to be in excess of 1,200 percent! Yet only 190 out of 1,070 plots were used for pineapple. When the authors asked farmers why they were not farming pineapple, the virtually unanimous response was, “I don’t have the money,”²¹ although some heterogeneity in ability between those who have switched to pineapple and those who have not, cannot be entirely ruled out.

Evidence from experimental farms suggests that, in Africa, the rate of returns to using chemical fertilizer (for maize) would also be high. But the evidence may not be realistic if the ideal conditions of an experimental farm cannot be reproduced on actual farms. Foster and Rosenzweig (1995) show, for example, that the returns to switching to high-yielding varieties were actually low in the early years of the green revolution in India, and the returns were even negative for farmers without an educa-

tion. This, despite the fact that these varieties had been selected precisely for having high yields, in proper conditions. But they required complementary inputs in the correct quantities and timing. If farmers were not able or did not know how to supply them, the rates of returns were actually low.

Chemical fertilizer, however, is not a new technology, and the proper way to use it is well understood. To estimate the rates of returns to using fertilizer on farms in Kenya, Duflo, Kremer, and Robinson (2004), in collaboration with a small non-governmental organization (NGO), set up small randomized trials on people's farms. Each farmer in the trial delimited two small plots. On one randomly selected plot, a field officer from the NGO helped the farmer apply fertilizer. Other than that, the farmers continued to farm as usual. The rates of return from using a small amount of fertilizer varied from 169 percent to 500 percent, depending on the year, although marginal returns declined quickly with the quantity of fertilizer used on a plot of a given size.

Evidence for a different type of underinvestment in agriculture is the negative size-productivity relationship, the idea that the smallest farms tend to be the most productive (table 5.2). The gap in the productivity of small and large farms within a country can be enormous: a factor of 5.6 in Brazil and a factor of 2.75 in Pakistan.²² It is smaller in Malaysia (1.5), but a large farm in Malaysia is not very large. This is strong *prima facie* evidence that markets are somehow not allocating the right amount of land to those who currently farm the smaller plots.

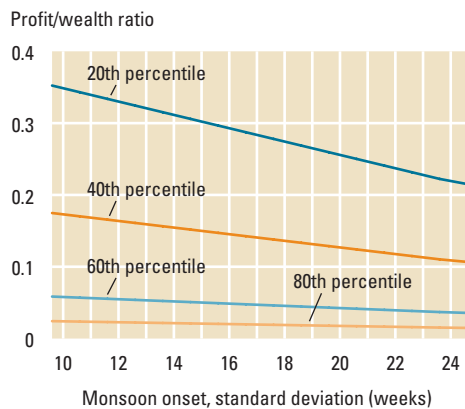
The problem with this kind of evidence is that it ignores the many reasons why the bigger farm may be inherently less productive, for example, lower soil quality. Even so, similar (but somewhat less dramatic) results show up even after controlling for differences in land quality. The profit-wealth ratio in Indian ICRISAT villages is the highest for the smallest farms, and when risk is comparatively low, the gap is more than 3:1 (figure 5.6). Because wealth includes the value of the land, the measure implicitly takes into account differences in the quality of the land. It does so as long as land prices are a

Table 5.2 Farm size productivity differences, selected countries

Farm size	Northeast Brazil	Punjab, Pakistan	Muda, Malaysia
Small farm (hectares)	563 (10.0–49.9)	274 (5.1–10.1)	148 (0.7–1.0)
Largest farm (hectares)	100 (500+)	100 (20+)	100 (5.7–11.3)

Source: Berry and Cline (1979).
Note: 100 = land productivity in the largest farm size.

Figure 5.6 Profit-wealth ratios are highest for the smallest farms



Source: Rosenzweig and Binswanger (1993).
Note: The standard deviation of the date of monsoon onset is a measure of underlying risk. The onset date of the monsoon was the single most powerful of eight rainfall characteristics to explain gross farm output. The data come from the Indian ICRISAT villages.

reasonable measure of land quality, which, however, is not entirely clear. There are also residual doubts about whether the returns are well measured—it is possible that the land of the smaller farms is degrading faster, but the degradation is not being counted while calculating the returns.

For these same firms, when risk goes up, the average return goes down. In part this may be inevitable, but it may also reflect the fact that the lack of insurance encourages people to avoid risky (but remunerative) choices.²³ This is consistent with the fact that profitability falls faster for the poorer farmers (less able to self-insure) as the risk goes up. Specifically, a one-standard-deviation increase in the coefficient of variation of rainfall leads to a 35 percent reduction in the profit of poor farmers, a 15 percent reduction in the profit of median farmers, and no reduction in the profit of rich farmers. The study also finds that input choices are affected by variability in rainfall, and in par-

ticular, poor farmers make less efficient input choices in a risky environment.

In related work, Morduch (1993) specifically investigated how the anticipation of a credit constraint affects the decision to invest in high-yielding variety seeds. Using a methodology inspired by Zeldes (1989), he splits the sample into two groups—one group of landholders expected to have the ability to smooth their consumption, and one group that owns little land, expected to be constrained. He finds that the more constrained group devotes a considerably smaller fraction of land to high-yielding variety seeds for rice and castor.

Another consequence of the lack of insurance is that it may lead households to use productive assets as buffer stocks and consumption smoothing devices, which would be a cause for inefficient investment. Rosenzweig and Wolpin (1993) argue that bullocks (an essential productive asset in agriculture) serve this purpose in rural India. They show, using ICRISAT data covering three villages in semiarid areas in India, that bullocks, which constitute a large part of households' liquid wealth (50 percent for the poorest farmers), are bought and sold quite frequently (86 percent of households had either bought or sold a bullock in the previous year). Moreover, they buy when they are flush with money and sell when they are broke.

Since people are not simultaneously selling and buying land, they are not selling these animals because they no longer need them for production. Indeed, from the view point of production, most of these farmers should own two bullocks and never sell them. If they are selling, the reason is that they need the money for consumption. The data suggest that, for poor or midsize farmers, there is considerable underinvestment in bullocks, presumably because of the borrowing constraints and the inability to borrow and accumulate financial assets to smooth consumption: almost half the households in any given year hold no bullocks (most of the others own exactly two).²⁴

There is also compelling evidence that sharecroppers lack incentives. Binswanger

and Rosenzweig (1986) and Shaban (1987) both show that productivity is 30 percent lower in sharecropped plots, controlling for farmers' fixed effects (that is, comparing the productivity of owner-cultivated and farmed land for farmers who cultivate both their own land and that of others) and for land characteristics. Shaban (1987) shows that all the inputs are lower on sharecropped land, including short-term investments (fertilizer and seeds). He also finds systematic differences in land quality (owner-cultivated land has a higher price per hectare), which could in part reflect long-term investment.

On the impact of security of property, Do and Iyer (2003) find that a land reform that gave farmers the right to sell, transfer, or inherit their land-use rights also increased agricultural investment, particularly the planting of multiyear crops (such as coffee). Laffont and Matoussi (1995) use data from Tunisia to show that a shift from sharecropping to owner cultivation raised output by 33 percent, and moving from a short-term tenancy contract to a longer-term contract increased output by 27.5 percent.²⁵

Security of property rights is often linked to the local power structure. The connection between inequalities in power and underinvestment is nicely exemplified by the Goldstein and Udry (2002) study of investment in land in a setting where land is allocated by custom (rural Ghana). They show that individuals are less likely to leave their land fallow (an investment in long-run productivity of the land) if they do not hold a position of power within either the hierarchy of the village or the hierarchy of the lineage. The problem is that the land gets taken away from them when it is lying fallow. Because women rarely hold these positions, women's land is not left fallow enough and is much less productive than men's.

Human capital

According to the report of the Commission on Macroeconomics and Health (2001), returns to investing in health are on the order of 500 percent. But these numbers, arrived at through cross-country growth

regressions, are not as easy to interpret as what would actually happen if someone were to invest an extra dollar in health. That said, there clearly are examples of specific health interventions that have enormous private and social returns. There is substantial experimental evidence that supplementation in iron and vitamin A increases productivity at relatively low cost.

- Basta, Soekirman, and Scrimshaw (1979) study iron supplementation among rubber tree tappers in Indonesia. Baseline health measures indicated that 45 percent of the study population was anemic. The intervention combined an iron supplement and an incentive (given to both treatment and control groups) to take the pill on time. Work productivity among those who got the treatment increased by 20 percent (or \$132 a year), at a cost per worker-year of \$0.50. Even taking into account the cost of the incentive (\$11 a year), the intervention suggests extremely high rates of returns.
- Thomas and others (2005) obtain lower but still high estimates in a larger experiment, also in Indonesia. They found that iron supplementation experiments in Indonesia reduced anemia, increased the probability of participating in the labor market, and increased earnings of self-employed workers. They estimate that, for self-employed males, the benefits of iron supplementation amount to \$40 per year, at a cost of \$6 per year.²⁶
- The cost-benefit analysis of a deworming program²⁷ in Kenya reports estimates of a similar order of magnitude. Taking into account externalities (because of the contagious nature of worms), the program led to an average increase in schooling of 0.14 years. Using a reasonable figure for the returns to a year of education, this additional schooling will lead to a benefit of \$30 over the life of the child, at a cost of \$0.49 per child per year. Not all interventions have the same rates of return, however. A study of Chinese cotton mill workers²⁸ led to a significant increase in fitness, but no corresponding increase in productivity.

Measured returns to private investment in education tend not to be quite so high. Banerjee and Duflo (2004b) survey cross-country evidence, and conclude that—

Using the preferred data, the Mincerian rates of returns seem to vary little across countries: the mean rate of returns is 8.96, with a standard deviation of 2.2. The maximum rate of returns to education (Pakistan) is 15.4 percent, and the minimum is 2.7 percent (Italy).²⁹

But most of the educational benefits of deworming mentioned above would be captured by a child whose parents are willing to spend \$0.50 on the deworming medicine. This clearly offers a return much higher than the measured Mincerian returns at affordable absolute cost, although they are not strictly comparable. Deworming does not require the child to spend more years in school, but it does help the child get more out of the years he or she is already spending in school. However, when the deworming medicine was offered free to the children, the take-up was only 57 percent. In this sense, it is clear that at least some causes of underinvestment have to be found in the way the family makes decisions, rather than in the lack of resources.

The fact that a lack of connections alters the nature of human capital investment is nicely demonstrated in a recent paper by Munshi and Rosenzweig (forthcoming). They show that, in India, trade liberalization increased returns to knowing the English language in families with connections in the blue-collar sector compared with families with no connections. However, there is a much bigger gap between girls and boys in the increase in enrollment in English-medium schools. This is attributed to the fact that girls never really expected to get these blue-collar jobs, while for their brothers, it depended on whether they had the right contacts.

Inequalities and investment

Four important points follow from this body of evidence: first, markets in developing countries are highly imperfect, and

those who do not have enough wealth or social status tend to underinvest. The resources underused because of this underinvestment end up being used for some less productive purpose, reducing overall productivity. In the example from the knitted garment industry in Tirupur, the Gounders were overinvesting in their own relatively unproductive firms, while the much more productive firms of the outsiders were starved of capital. The land owned by Ghanaian women was getting degraded, because they did not have the social status needed to hold on to the land during the following period. This, once again, is a pure loss for society. The fact that other people who do have status and can fallow their land as needed is not, in any way, compensating for the loss of productivity on the lands of the powerless. This creates a strong presumption that certain specific types of redistribution, by empowering certain people or increasing their access to resources or contacts, can promote efficiency and equity.

Second, this hypothesis would imply a bias in favor of those kinds of redistribution that target the specific lack of access to resources or influence causing the inefficiency. In some situations this will mean redistributing assets, but it also might mean redistributing access to capital, perhaps by promoting microcredit, strengthening women's land rights or access to jobs and welfare programs, designing affirmative action programs to break down stereotyping, and improving access to justice systems.

Third, because investments build wealth and wealth makes it easier to invest in a world where markets do not function very well, a little help can go a long way. Starting the right business might be the biggest challenge: once started, the business might propel itself forward without any further help.

Fourth, it is not clear that the beneficiaries from this kind of efficiency-promoting redistribution have to be the poorest of the poor. Because the ideal is to promote productive investments, the target should be those most likely to make these investments. Whether the poorest are the right people from this point of view is an empirical question, and one for which the answer

might depend on the set of economic opportunities available.

The microcredit community, in particular, has long debated this last issue in trying to decide whether microcredit is best instrument for helping the poorest of the poor. This clearly turns partly on whether the poorest are the ones who have the projects with the highest returns, which could be the case if the poor and the less poor have the same kinds of production functions, and if there are diminishing returns to scale. If, instead, the most productive technology in this area had a fixed cost of production but (say) diminishing returns otherwise, giving the poorest access to more capital may not be very productive: even with all the capital they can get, they may not be able to cover the fixed cost. It may be more effective to help people who are slightly richer, because with some help they may actually be able to start a business.

How good or bad is the assumption of decreasing returns in the production function of an individual firm? As mentioned above, McKenzie and Woodruff (2003) estimate a production function for small Mexican firms, suggesting strong diminishing returns. Mesnard and Ravallion (2004) find weak diminishing returns using Tunisian data. But estimating a production function that exhibits local increasing returns is inherently difficult. A firm is likely to grow (or shrink) quickly when it is in the region of increasing returns. So we will observe few firms in this region, and be likely to reject too often the assumption of local increasing returns. Certainly the natural interpretation of the results in Banerjee and Duflo (2004a), showing close to 100 percent returns in medium firms in India, is that there are increasing returns over some range.

A corollary of this discussion is that the redistribution that maximizes productivity growth is not necessarily the one that has the strongest immediate effect on poverty. Nor is it the one that does most to reduce inequality. Indeed, except under very special circumstances, this discussion tells us nothing about the relation between some global measure of inequality and the efficiency of resource use or investment. Consider the case, discussed above, in which the

production function has a fixed cost but also diminishing returns. If all firms are equal and the maximum they can each invest is less than the fixed cost, no one will be able to start a firm. Increasing inequality will raise the productivity of capital by making it possible for some firms to pay the fixed cost. Because there are also diminishing returns, however, there will be a point at which any further increase in overall inequality would be counterproductive.

More generally, the effect of inequality will depend on the shape of the production function, and the size of the investment potential of the average person relative to the fixed cost. Obviously, the issue gets even more complicated if different firms have different production functions and if productivity is correlated with the owner's wealth (as it might be if the owner's education is an important input into production and richer people tend to be more educated).

Several authors have tried to look for a systematic relation in cross-country data between inequality and growth (presumably what investment is meant to achieve). A lengthy body of literature³⁰ estimated a long-run equation, with growth between 1990 and 1960, for example, regressed on income in 1960, a set of control variables, and inequality in 1960. Estimating these equations tended to generate negative coefficients for inequality. But there are obvious concerns about whether such a relation could be driven entirely by omitted variables. To address this problem, Li and Zou (1998), Forbes (2000), and others used the time series dimension of the Deininger and Squire data set to look (effectively) at the effect of short-run changes in inequality on changes in growth.³¹ The results change rather dramatically: the coefficient of inequality in this specification is positive and significant.

A recent review paper by Voitchovsky (2004) concludes that both these effects are quite robust. Most studies that look at the cross-sectional relationship between inequality and subsequent growth over a relatively long period in cross-country data, and especially those that use measures of asset inequality, find a negative relationship, often significant.³² By contrast, most studies that look at the relationship between changes in

inequality and changes in growth, including several studies that do the analysis at the sub-national level within the same country, find a positive effect.

Both Banerjee and Duflo (2003) and Voitchovsky (2004) conclude that there is no reason to give one of these sets of results priority over the other. Indeed, both could be right. For example, in the short run, policies that allow large cuts in real wages might encourage investment, but in the long run, the consequent increase in poverty might make it harder for the population to maintain its human capital. Or both could be wrong. Most important among the many reasons for both the cross-sectional and the time series evidence to be misleading are the following: the possibility of a nonlinear relationship between inequality and growth, problems with comparability of cross-country data, and the difficult question of identifying the direction of causality when both variables are likely to influence one another.

This lack of clear-cut results is perhaps disappointing, but it is worth emphasizing that our focus here has been on redressing specific inequalities in productive opportunities rather than some overall measure of inequality. Despite the great attention devoted to the question of a systematic relationship between overall inequality and growth at the country level, the body of evidence remains unconvincing. But there clearly are situations in which there is a strong presumption that reducing a specific inequality would promote better investment.

One such example comes from *Operation Barga*, a tenancy reform in the Indian state of West Bengal in the late 1970s and 1980s. It has been known, at least since the work of the great Victorian economist Alfred Marshall, that sharecropping provides poor incentives and discourages effort. In such an environment, a government intervention that forces the landlords to give their sharecroppers a higher share of the output than the market would give them should increase effort and productivity. This is exactly what happened in West Bengal, India, when a Left Front government came to power in 1977. The tenant's share of output was set at a minimum of 75 percent as long as the tenant provided all

inputs. In addition, the tenant was guaranteed a large measure of security of tenure, which may have encouraged him or her to undertake more long-term investments on the land. Survey evidence shows a substantial increase in both the security of tenure and the share of output going to the sharecropper. The fact that the implementation of this reform was bureaucratically driven, and proceeded at different speeds in different areas, suggests the possibility of using variation in the implementation of the reform to evaluate its impact. The evidence suggests that there was a 62 percent increase in the productivity of the land.³³

A different program, also promoting equity and efficiency, had to do with redressing the effects of intrafamily inequality. A long line of research claims that income and expenditures are often controlled by the male members of the family and that this leads to underinvestment, especially in the health and education of girls. One fallout of dismantling the apartheid regime in South Africa was the expansion of the South African social pension program to the black population. Pension entitlements would accrue to elderly males and females, and many older women living alone were entitled to receive the benefit. In many cases, children of very poor parents were sent to live with grandparents who began to receive these pensions. Duflo (2003) compared the impact of these new transfers on the nutrition of children living with their grandparents, separately for households in which the pension was given to the grandmother and those in which it was assigned to a grandfather.

For children born before the expansion, in 1990 and 1991, height-for-age was slightly lower in families in which the grandmother would eventually get the pension. For children born after the expansion, in 1992 and 1993, the children are significantly taller (except for the newborns) in those families. There is no difference between noneligible families and families in which pension

money goes to the grandfather. (Boys are essentially unaffected.) The estimates suggest that receipt of the pension (which was about twice the per capita income among blacks) was enough to help girls bridge half the gap in height-for-age between South African and American children.

These examples show that it is possible to enhance both equity and efficiency simultaneously. Judicious redistribution—of income to grandmothers, of power to poor women farmers, of credit to entrepreneurs in small firms—can increase the productivity of resources, such as land, human capital, and physical capital. If markets fail, resources do not always flow to where their return is greatest, particularly if that happens to be in projects run by people with limited wealth or influence. Careful microeconomic case study evidence, some of which was summarized in this chapter, suggests that certain forms of redistribution can reduce waste and contribute to a better use of resources, while also reducing inequality of opportunity. In fact, it enhances efficiency precisely because it reduces inequality of opportunity.

This is not to say that one cannot easily imagine certain types of redistribution that hurt efficiency. But given the near universality of market failures and underinvestment in poor countries, it should be possible, with a combination of good research and careful thinking, to identify opportunities for redirecting resources to poorer people who are in a position to make good use of them.

In making the case for improvements in equity that are also efficiency-enhancing, this chapter used mainly microeconomic evidence on markets, wealth, and agency of individuals. The next chapter uses a different set of historical, macroeconomic, and institutional evidence to argue that complex historical processes, combined with inequalities in influence and power, may lead to bad political and economic institutions, which severely impair the development of poor countries.

Equity and development in the Spanish transition to democracy

In the last half century, Spain has gone from authoritarianism and underdevelopment to democracy and wealth. Spain's history illustrates how the distribution of political agency and economic assets greatly influences the policy choices available to a society. The fundamentals of economic and political structure influence and constrain the choices. But the process is not deterministic: political agency and policies can shift the underlying fundamentals (as happened in Spain in the 1960s and 1970s) and open the space for new choices.

Before the civil war: social and economic polarization

Until the second half of the twentieth century, Spanish contemporary history was a tale of political and economic failure. After a period of territorial expansion and European hegemony in the early modern ages, Spain lapsed into economic decline and cultural stagnation in the following centuries. During most of the nineteenth century, its industrial takeoff was blocked by political instability, inefficient legal institutions, substantial inequalities, and a poorly educated population. In 1929, per capita income was \$3,000 (in 1990 dollars)—two-fifths that of Britain and less than two-thirds that of France.

Spain was polarized by entrenched social and economic inequalities. In a country still eminently agrarian, the distribution of land was very unequal. About 1 percent of the holdings occupied 50 percent of the land. Educational attainment remained low, strictly linked to circumstances of birth. Social mobility was almost nonexistent. Except for Catalonia and the Basque country, which industrialized in the nineteenth century, Spain lacked a large middle class.

Against this backdrop of relative stagnation and high inequality, democratic institutions were introduced in 1931—only the second time in Spanish history. They did not last long. The brief democratic period (1931–36) was characterized by huge political instability and social agitation. The first Republican government pushed ahead with a strong reformist program: separation of church and state; a single system of state schools and a goal to universalize education; a process of land reform; a law to decentralize political power to Catalonia; and stepped up efforts to reform the army.

These reforms elicited a strong reaction from the right, which came to power in 1933 and quickly moved to halt them. Two and a half years later, in the spring of 1936,

new elections were held, with a victory for the left. The threat of more radical policies prompted a military uprising, supported by the landed classes, much of the bourgeoisie, and the church. Spain became engulfed in a three-year civil war. The social polarization locked Spain into a zero-sum game over the distribution of wealth. There was little political space for compromise or reformist solutions. The “haves” opposed all attempts at even minimalist reform. The “have-nots” wanted radical change, not gradual reform.

The Franco regime: from autarky to growth

With the defeat of the Republican government by the Nationalist army in 1939, Generalissimo Francisco Franco established an authoritarian regime that lasted until his death in 1975. The destruction caused by the civil war depressed the Spanish economy. Per capita income fell to its 1900 level and did not reach its 1918 level until 1950. The proportion of the active population in industry declined to 22 percent in 1940 (the level in 1920) and the share of employed in agriculture rose above 50 percent. Growth averaged only 1.2 percent a year in the 1940s.

Spain's economic recovery was hampered, above all, by the autarkic and statist policies of the Franco regime. Inspired by the corporatist ideologies of Italian Fascism and German Nazism, Franco's regime generalized a system of price controls and rationing and regulated foreign trade through quantitative controls. This interventionist strategy extended to the labor and housing market. To quell one of the main forces that opposed the military insurrection, Franco outlawed any independent labor unions. Instead, workers and employers had to affiliate in a national trade union organization. This repressive stance was “compensated” by strict labor legislation that made it hard for employers to dismiss workers or to hire them through temporary

contracts. Emphasis on permanent jobs and cheap housing was seen as a substitute for the lack of direct social policies, an attempt by the regime to win legitimacy.

In the late 1950s, Spain eventually moved to break with this interventionist system. An acute political crisis—associated with a wave of strikes, an economic recession, and severe balance-of-payment crisis—led the government to adopt a stabilization plan in March 1959. In addition to fiscal and monetary restraint, the plan included wide-ranging measures to liberalize the economy. It was an outright success. From 1960 to the outbreak of the first oil crisis, output expanded at an average annual rate of more than 7 percent with very little interyear volatility. Per capita income almost tripled from about \$3,000 (in 1990 dollars) to \$8,500 in 15 years. Productivity growth averaged 6 percent.

The transformation of the Spanish economy led to significant structural changes in Spanish society. The combination of economic growth, industrial expansion and internal migration produced a substantial decline in the levels of interregional inequality (from a standard deviation in per capita income of 0.37 in 1955 to 0.27 in 1973). Interhousehold inequality also declined considerably: the Gini coefficient for wages and salaries of employees (agrarian and industrial) declined from 0.29 in 1964 to 0.23 in 1973; the Gini coefficient for household income fell from 0.39 in 1964 to 0.36 in 1974. The income share of the three central deciles went up from about 51 percent to 59 percent in that decade.

Still, significant social and economic inequities remained. Although the illiteracy rate had fallen to 10 percent by 1970, only 6 percent of the population had completed secondary studies. Wages remained dampened by repressive labor institutions. Taxation and public spending were low, and redistributive social programs nonexistent.

Transitioning to democracy and building the welfare state

Following Franco's death in 1975, King Juan Carlos became the Spanish head of state. He immediately launched a process of political change. Employing the legal mechanisms put in place by the very technocratic generation that had reformed the economy in the early 1960s, as well as pointing to wide popular support for democracy, he secured the consent of the old Francoist *Cortes* to establish a truly democratic parliament elected through direct, competitive elections.

The political reform was ratified with overwhelming popular support in a referendum in December 1976. Although conducted in a climate of uncertainty, particularly over the reaction of the army and the extent to which terrorist violence or labor mobilization could disrupt the negotiations, democratic elections were held in June 1977. After protracted negotiations, a new constitution was approved in 1978 with the support of all parliamentary groups. To reinforce the political pact in parliament, the government also struck a wide economic and social deal with employers and trade unions that same year.

Spain's democratization was rooted in the new economic and social conditions of

the 1960s and 1970s. Rapid industrialization and urbanization deflated past conflicts around the distribution of land. The expansion of literacy and the increase in productivity and incomes generated a large middle class. Sustained growth defused social conflict with the credible promise of higher incomes and more social mobility. In short, Spain had overcome the zero-sum game it had been locked in for the past century and a half.

Economic growth resulted in a different economic structure and better distributional outcomes, supporting a swift and successful transition to democracy. In turn, the transition to democracy changed the role and size of the public sector.

Democratization reinforced social demands for progressive and redistributive policies—especially for public infrastructure, and education, health, and social programs. In 1979 more than 70 percent of Spaniards agreed with the statement that “the distribution of wealth in this country is totally unjust.” In 10 years, social expenditure almost doubled to reach 80 percent of the European average. Public expenditure in education steadily increased from 2 percent of GDP in 1975 to 4.5 percent in 1995. By 2001 almost 50 percent of the population had completed

secondary education—10 times more than in the mid-1970s. An ambitious public investment program tripled the public highway network, revamped and expanded metropolitan transportation, and modernized the railroad system.

Spain's transition to democracy and the resulting expansion of its welfare state shows how a mutually reinforcing package of policy and institutional choices leading to greater equity helped underpin the development and modernization of the Spanish economy and its integration into the European Union. It illustrates how political and economic structures shape the possibilities for policy choice, a theme of chapter 6. But it also illustrates that specific policy choices matter—across social sectors, infrastructure, the workings of markets, and international integration—and that there can be important complementarities for both equity and dynamic growth, notably between greater social provisioning and greater reliance on markets. This takes us to the issue of practical policy design, the central theme of part III of this report.

Sources: Synthesized from Boix (2005), with references to Gunther, Montero, and Botella (2004); North and Thomas (1973); and Revenga (1991).