Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Contracting Constraints, Credit Markets and Economic Development

Abhijit Banerjee

Working Paper 02-17
September 2001

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
http://papers.ssrn.com/paper.taf?abstract_id=306990

Contracting Constraints, Credit Markets and Economic Development

Abhijit V. Banerjee*
Department of Economics, MIT

September, 2001

Abstract

This paper begins by summarizing the micro-evidence on credit markets from a large number of studies from all over the world, with the goal of identifying a number of stylized facts. We argue that, in particular, the evidence strongly suggests that for poor people in developing countries, imperfections in the credit market are quantitatively very important. We then build a simple model that explains the observed patterns, based on the idea that monitoring and screening borrowers have both fixed and variable costs. We go on to build a simple dynamic model that allows us to understand what the observations about the credit market imply for the evolution of the wealth distribution.

JEL Codes: O12, D82, D31
Keywords: Credit Markets; Distribution; Growth

1 Introduction

Development economists are, perhaps by necessity, optimistic people. One does not become a development economist if one believes that the world's poorest are doing as well as they possibly could. Indeed the premise of the entire field[1] is that there is talent in every people, if not every person, and if

*This paper grew out a lecture I gave at the World Congress of the Econometric Society in August 2000. I am grateful to Philippe Aghion and Esther Duflo for their comments and Marko Tervio for his help with simulations.

[1] And of growth theory: What is convergence other than the hope that there is talent in every nation?

there is one central question it has to be what prevents people from making the best use of their natural talents?

There are at least five distinct answers to this question. The first, which is elaborated here, is the answer from contract theory: Talent is not an apple,[2] one cannot simply go to the market, sell one's talent and expect to be paid the appropriate price. The second is coordination failure: Talent is only talent if it gets to work with the appropriate other inputs. Even Lennon needed Paul and George—had they decided to go to the City instead, he too might have found himself a different profession.[3] The third is political economy: Governments can and often do make it harder for people to do what they are best at doing.[4] The fourth is learning: People may not know what they ought to be doing, and even when they do the rest of the world may not appreciate them. For example, a growing body of evidence shows that farmers are often ignorant or suspicious of more rewarding crops and better seeds.[5] The final answer comes from what has come to be called behavioral economics: People may not always seek out the best options because they are held back by psychological constraints or social norms.[6]

The fact that this survey concentrates on the contract theoretic argument should not be interpreted as evidence for its primacy. But it is the argument that has received the most elaboration over the last decade or so, and the one that best matches the competencies of the present author. It is therefore the appropriate topic for a survey like this.

[2] A contract theorist would probably say an apple is not an apple either—they can be stale or fresh, sweet or sour. If you had a lot of apples to sell, you would probably want to invest in a reputation for selling only fresh and sweet apples.

[3] There is a long tradition in development economics of models that emphasize coordination failures going back at least to Paul Rosenstein-Rodan (1943) and Ragnar Nurkse (1953). See Murphy, Shleifer and Vishny (1989) for a model where there is failure of coordination between producers of different goods, and Kremer (1993) for a model for a model of matching between different types of talents.

[4] There is, of course, a long tradition here, going back to Adam Smith. Krueger (1974) and Bhagwati (1982), among others, study the distortions in the allocation of talent and resources that come from government policies.

[5] See Besley and Case (1994) and Munshi (2000). Banerjee (1992) provides theoretical arguments for why such behavior may be rational for individual farmers.

[6] There is a long and controversial literature on this point. The famous Lewis model (Lewis, 1954) argued that family norms could discourage people from seeking outside options. The rational peasant model (Schultz, 1964) was articulated as a critique of models like the Lewis model (see also Cole, Mailath and Postlewaite, 1992, and Banerjee and Newman, 1998, for two very different attempts to reconcile these views). It is, however, time to revisit this issue: With the increasing sophistication of the psychological models used in economics, it is now possible to re-ask the question of whether, for example, poverty can have direct discouragement effects.

2 Contract Theory in Development Economics

Contract theoretic arguments in development economics go back at least to the work of D. Gale Johnson in the 1940's and 1950's in the context of land markets.[7] Stiglitz's 1974 paper on sharecropping, among others, started a tradition of formal contract theoretic models that seek to explain why landlords and tenants often settle into arrangements that are, at least apparently, less than first best efficient.[8]

Since then, similar principles have been applied to the study of all the other important markets: capital, insurance and human capital. The result is an enormous literature which I could not even begin to do justice to within the limits of this survey. I will confine myself, therefore, to elaborating on a single example from the market for capital, which I hope will allow me to draw out the most important themes, though at several points in the text I will point out the connections with what is understood about the other asset markets.

3 The Credit Market

The facts about the credit market are remarkably stark. While neoclassical theory predicts a single price of capital at which people both borrow and lend, at any point over the last twenty years one could point to a set of peoples in world (most recently the Japanese) who were earning a negative return on their savings, while another set of people were borrowing at real rates of 60% or more.

Indeed, more often than not, very large differences between borrowing and lending rates can be found within a single sub-economy. Banerjee (2001) reviews a number of empirical studies of individual credit markets in developing countries and lists six salient features:[9]

1. **Sizeable gap between lending rates and deposit rates within the same sub-economy:**

[7] See Johnson (1950).

[8] See Stigltiz (1974) and Cheung (1968).

[9] The review focuses on the informal sector because the formal banking sector in most developing countries has tended to be quite rigid (interest rate caps, strict rules about collateral, inflexible credit limits, etc. [see Ghate, 1992]) with the result that the informal sector became the supplier of the marginal units of capital for all but the very largest firms.

Ghatak (1976) reports data on interest rates paid by cultivators in India from the *All India Rural Credit Survey* for the 1951-2 to 1961-2 period: The average rate varies between a maximum of 18% (in 1959-60) and a minimum of about 15% (in 1961-62). These numbers are, however, slightly misleading: around 25% of the borrowing reported in these surveys were zero-interest loans, usually from family members or friends. These should be seen as gifts/insurance rather than loans. If these were left out, the average rates in these surveys would be above 20%. We are not told what the comparable rates for depositors were in this period, but Ghatak reports that the bond rate in this period was around 3% and the bank deposit rate was probably about the same.

Timberg and Aiyar (1984) report data on indigenous style bankers in India, based on surveys that they carried out. They report the gap between the average rate charged to borrowers and the average rate to depositors by Finance Companies was 16.5%. The same gap for financiers from the Shikarpuri community was 16.5%, 12% for financiers from the Gujerati community, 15.5% for the Chettiars, 11.5% for the Rastogis, etc.

The "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989) reports results from a number of case studies that were commissioned by the Asian Development Bank and carried out under the aegis of the National Institute of Public Finance and Policy. For the rural sector, the data is based on surveys of 6 villages in Kerala and Tamil Nadu, carried by the Centre for Development Studies. The average interest rate charged by professional money-lenders (who provide 45.61% of the credit) in these surveys is about 52%. While the average deposit rate is not reported, the maximum from all the case studies is 24% and the maximum in four out of the eight case studies is no more than 14%. For the urban sector, the data is based on various case surveys of specific classes of informal lenders: For Finance Corporations they report that the maximum deposit rate for loans of less than a year is 12% while the minimum lending rate is 48%. For hire-purchase companies in Delhi, the deposit rate was 14% and the lending rate was at least 28%. For auto-financiers in Namakkal, the gap between the deposit rate and the lending rate was 19%.[10] For handloom financiers in Bangalore and Karur,

[10] This number and all other information about this gap is measured in percentage

the gap between the deposit rate and the lowest lending rate was 26%.[11]

Aleem (1990) reports data from a study of professional moneylenders that he carried out in a semi-urban setting in Pakistan in 1980-1981. The average interest rate charged by these lenders is 78.5%. The bank rate in that year in Pakistan was 10%. However, it is possible that depositors in this area may not have been depositing in the banks, so an alternative measure of the gap can be obtained by using the Aleem's numbers for the opportunity cost of capital to these money-lenders, which is 32.5%.[12]

2. **Extreme variability in the interest rate charged by lenders for superficially similar loan transactions within the same economy:**

Timberg and Aiyar (1984) report that the rates for Shikarpuri financiers varied between 21% and 37% on loans to members of local Shikarpuri associations and between 21% and 120% on loans to non-members (25% of the loans were to non-members and another 50% were loans through brokers). On the other hand, the Gujerati bankers charged rates of no more than 18%. Moreover, the rates faced by established commodity traders in the Calcutta and Bombay markets were never above 18% and could be as low as 9%.

The "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989) reports that Finance Corporations offer advances for a year or less at rates between 48% per year and the utterly astronomical rate of 5% per day. The rates on loans of more than a year varied between 24% and 48%. Hire-purchase contracts offer rates between 28% to 41% per year. Handloom Financiers charge rates between 44% and 68%. Yet the Shroffs of Western India offer loans at less than 21% and Chit Fund members can borrow at less than 25%.

points.

[11] A number of other lending instituions are also mentioned in this study. However, the range of both deposit rates and lending rates is so wide in these cases that the gap between the minimum lending rate and the maximum deposit rate is not very large. This does not rule out the possibility that the gap between the *average* borowing and lending rate is quite substantial even in these cases.

[12] This, however, understates the gap, since the money-lenders themselves borrow this money and the original lenders are paid much less than 32.5%.

The same report tells us that among rural lenders, the average rate for professional money-lenders (who in this sample give about 75% of the commercial informal loans) was 51.86%, whereas the rates for the agricultural money-lenders (farmers who also lend money) who supply the rest was 29.45%. Within the category of professional money-lenders, about half the loans were at rates of 60% or more but another 40% or so had rates below 36%.

The study by Aleem (1990) reports that the standard deviation of the interest rate was 38.14% compared to an average lending rate of 78.5%. In other words, an interest rate of 2% and an interest rate of 150% are both within two standard deviations of the mean.

Swaminathan (1991) reports on a survey of two villages in South India that she carried out: The average rate of interest in one village varied between 14.8% for loans collateralized by immovable assets (land, etc.) and 60% for loans backed by moveable assets. The corresponding rates in the other village were 21% and 70.6 %. Even among loans collateralized by the same asset—gold—the average rate in one village was 21.8% but it went up to 58.8% when the loans were to landless laborers.

Ghate (1992) reports on a number of case studies from all over Asia: The case study from Thailand found that interest rates were 2-3% per month in the Central Plain but 5-7% in the north and north-east (note that 5 and 7 are very different).

Gill and Singh (1997) report on a survey of 6 Punjab villages they carried out. The mean interest rate for loans up to Rs 10,000 is 35.81% for landowning households in their sample, but 80.57% for landless laborers.

Fafchamps' (2000) study of informal trade credit in Kenya and Zimbabwe reports an average monthly interest rate of 2.5% (corresponding to annualized rate of 34%) but also notes that this is the rate for the dominant trading group (Indians in Kenya, whites in Zimbabwe). Blacks pay 5% per month in both places.[13]

Irfan et al. (1999), mentioned above, report that interest rates charged by professional money-lenders vary between 48% and 120%.

3. **Low levels of default:**

[13] Fafchamps notes that when he controls for the sector of the economy, etc., this difference goes away but that just tells us that the source of the variation is sector rather than race.

Timberg and Aiyar (1984) report that average default losses for the informal lenders they studied ranges between 0.5% and 1.5% of working funds.

The "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989) attempts to decompose the observed interest rates into their various components,[14] and finds that the default costs explain 14 per cent (not 14 percentage points!) of the total interest costs for the Shroffs, around 7% for auto-financiers in Namakkal and handloom financiers in Bangalore and Karur, 4% for Finance Companies, 3% for hire-purchase companies and essentially nothing for the Nidhis.

The same study reports that in four case studies of money-lenders in rural India they found default rates explained about 23% of the observed interest rate.

The study by Aleem gives default rates for each individual lender. The median default rate is between 1.5 and 2% and the maximum is 10%.

4. **Production and trade finance are the main reasons given for borrowing, even in cases where the rate of interest is relatively high:**

Ghatak (1976) concludes on the basis of his study that "the existing belief about the unproductive use of loans by Indian cultivators ... has not been substantiated."

Timberg and Aiyar (1984) report that for Shikarpuri bankers (who charge 31.5% on average, and as much as 120% on occasion), at least 75% of the money goes to finance trade and, to lesser extent, industry.

The "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989), reports that several of the categories of lenders that have been already mentioned, such as hire-purchase financiers (interest rates between 28%-41%), handloom financiers (44%-68%), Shroffs (18%-21%) and Finance Corporations (24%-48% for longer term loans and more than 48% on loans of less than a year) focus almost exclusively on financing trade and industry, and even for Chit Funds and Nidhis, which do finance consumption, trade and industry dominate.

[14] In the tradition of Bottomley (1963).

Swaminathan (1991) reports that in the two villages she surveys, the share of production loans in the portfolio of lenders is 48.5% and 62.8%. The higher share of production loans is in Gokalipuram, which has the higher interest rates (above 36% for all except the richest group of borrowers).

Ghate (1992) also concludes that the bulk of informal credit goes to finance trade and production.

Murshid (1992) studies Dhaner Upore loans in Bangladesh (you get some amount in rice now and repay some amount in rice later) and argues that most loans in his sample are production loans despite the fact that the interest rate is 40% for a 3-5 month loan period.

Gill and Singh (1997) report that the bulk (63.03%) of borrowing from the informal sector goes to finance production. This proportion is lower for the landless laborers but it is an non-negligible fraction (36%).

5. **Richer people borrow more and pay lower rates of interest.**

Ghatak (1976) correlates asset category with borrowing/debt in the *All India Rural Credit Survey* data and finds a strong positive relationship.

Timberg and Aiyar (1984) report that some of the Shikarpuri and Rastogi lenders set a credit limit that is proportional to the borrower's net worth: Several lenders said that they would lend no more than 25% of the borrower's net worth, though another said he would lend up to 33%.

The "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989) tells us that in their rural sample, landless laborers paid much higher rates (ranging from 28-125%) than cultivators (who paid between 21 and 40%). Moreover, Table 15.9 in that report clearly shows that the average interest rate declines with loan size (from a maximum of 44% to a minimum of 24%). The relation between asset category and interest rate paid is less clear in their data but it remains that the second poorest group (those with assets in the range Rs 5,000-10,000) pays the highest average rate (120%) and the richest (those with more than Rs 100,000) pay the lowest rate (24%).

Swaminathan (1991) finds a strong negative relation between the value of the borrower's land assets and the interest rate he faces: The poorest (those with no land assets) pay 44.9% in one village and 45.4% in the other, while the rich (those with land valued at more than Rs 50,000) pay 16.9% and 24.2% in the corresponding villages.

Gill and Singh (1997) show that correlation between the wealth of the borrower and loan size is negative after controlling for the interest rate. They also find a positive relation between the borrower's wealth and the loan he gets.

6. **Bigger loans are associated with lower interest rates.**

Table 15.9 in the "Summary Report on Informal Credit Markets in India" (Dasgupta, 1989) clearly shows that the average interest rate declines with loan size (from a maximum of 44% to a minimum of 24%).

Ghate (1992) notes that the interest rate on very small loans in Bangladesh tends to be very high (Taka 10 per week on a loan of Taka 500, or 86% per annum).

Gill and Singh (1997) show that the correlation between loan size and the interest rate is negative even after they control for the wealth of the borrower.

3.1 Taking Stock: The Facts about Credit Markets

The fact that there is a gap between the lending rate and the rate paid to depositors is not, per se, surprising. The fact that intermediation is costly is, after all, entirely commonplace. What is striking is the size of the gap. It is always more than 10% and usually more than 14%, in a world where interest rates paid to depositors are rarely more than 20% and usually closer to 10%. In other words, intermediation costs seem to eat up at least a third and often half (and sometimes much more than half) of the income that could go to depositors.

However, this argument overstates the point slightly. The probability that a moneylender would default on his deposit liabilities is substantially lower than the probability that borrowers will default on the loan, which implies that the default premium on loans should be much greater than the default premium on deposits. From the evidence reported above, default is relatively rare and default costs rarely raise the interest rate by more than

10%. The gap between loan rates and deposit rates would be very large even if we were to deduct 10% from the loan rate.[15][16]

The fact that interest rates vary quite so much is particularly striking given the standard neoclassical prediction that in market equilibrium the marginal unit of capital in every firm should earn the same return. However, given that people might be rationed in the credit market, it is theoretically possible that the marginal product of capital is actually equal in all its uses, despite the enormous disparities in the interest rate. Note that the incremental capital/output ratio for the Indian economy is estimated to be around 4.3, implying an marginal return on capital of 24%. This is, however, a gross measure and the true return, net of depreciation, is clearly substantially lower (no more than 20%). The fact that interest rates above 35% are standard, and those above 75% are by no means rare, suggests that at least some of the users of capital must value capital at substantially more than 20%.

Could it be that all of the demand at relatively high interest rates comes from people who have particularly insistent consumption needs today? This is certainly not the stated purpose of the loans, as noted above. Of course, money is fungible and one cannot rule out the possibility that some of these people are either deluded or untruthful. However, it remains that when a handloom producer borrows at 48% or more to finance consumption, he chose to do so instead of taking the money out of his existing business. Therefore, it must be that he would lose more than 48% on any money that comes out of his business. This may be in part because in the short run his assets are not liquid, but this could not explain why he accepts ongoing financing on these terms. Therefore, he must be earning marginal returns that are close to 48%.[17]

[15] One might be worried that while default rates are low on average, default may be very important in those cases where the interest rate is high. However, this is not a problem since, for the most part, we look at interest rates and default rates weighted by volume (or equivalently, do a Bottomley [1975] decomposition). Moreover, in the one detailed micro-study we have where the average interest rate is very high (Aleem, 1990), default rates are actually very low (always less than 10%, and usually less than 2%).

[16] Delay in repayment for which no extra interest is charged is another factor that raises the interest rate. In Aleem's data, delay is much more common than default, but in a significant fraction of the cases the lender is able to charge interest for the extra days. Moreover, the percentage of loans that are late never exceeds 25% and the average delay is no more than 6 months, so at worst this would raise the interest rate by a factor of 1.12.

[17] There is, once again, the possibility that a part of the reason why these rates are so high is because of default risk. In other words, the expected return on the marginal units of capital need not be as high as 48%. However as already observed, defaults contribute relatively little to the level of the interest rate.

The fact that the marginal product of capital varies substantially across borrowers in the same sub-economy is supported by more direct evidence from the knitted garment industry based on data that I collected in joint work with Kaivan Munshi (Banerjee and Munshi, 2001). Tirupur produces 70% of India's knitted garment exports, and India is a major exporter of knitted garments. There are two communities of producers in Tirupur: Gounders, who are linked by community ties to a rich local agricultural community; and Outsiders, a motley crew of businessmen from all over India. They produce exactly the same goods, yet they use radically different technologies. Gounders invest much more than Outsiders at all levels of experience, both in absolute terms and relative to output. Average capital-output ratios for Gounders can be three times as large as that for Outsiders and is typically twice as large. However, all the evidence points to the Outsiders being more able: they enjoy faster output growth and their output outstrips that of the Gounders after a few years.

One possible situation where high ability people may invest less would be if capital were less useful for them, which would be the case if ability and capital were substitutes in the production function. The evidence, however, points against this explanation: When we compare Gounders with Gounders or Outsiders with Outsiders, it is clear that those who grow faster and produce more also invest more. Therefore, it seems relatively clear that the Outsiders invest less *despite having a higher marginal product of capital.*[18]

What explains why the credit markets behave in this way? Why is intermediation so inefficient with some people and so efficient with others? Why are the rich borrowers and those who borrow more favored by the market?

The standard theory of interest rates decomposes them into default rates, opportunity cost, transaction costs and monopoly rents. This is useful descriptively but stops well short of an explanation—the problem is that none of these can be seen as independent causal factors. Take the example of default rates. The fact that default rates are relatively low is not a fact about the nature of default: Timberg and Aiyar (1984) observe that some

[18] This is what Caballero and Hammour (2000) call scrambling. The proximate reason, it appears, is the Gounders have a lot of investible funds that they cannot profitably lend out because intermediation is so inefficient in India. Instead, they set up their own garment firms or lend to friends and family in the garment business. Since these firms are set up as a conduit for this surplus capital, they are not required to be particularly productive. The Outsiders, by contrast, come from traditional entrepreneurial communities and, as a result, their capital probably has many alternative uses. In other words, they do not invest in Tirupur because they lack other choices. This makes them more likely to be productive but also less willing to invest a lot.

branches of the state-owned commercial banks in India have default rates up to 60-70%. The low default rates observed in the studies we mention are a result of the steps taken by lenders to avoid default.

Monitoring the borrower is an obvious example of the kind of steps that lenders take. It is also an important source of what goies under the rubric of transaction costs: Aleem (1990) and Irfan et al. (1999) provide a list of steps taken by the lender to avoid default. These include getting to know the borrower through other transactions, visiting his establishment, making enquiries about him and going after him to make sure he repays.

Lenders also protect themselves by limiting their lending to borrowers they know.[19] This has four important consequences. First, it pushes capital towards well-connected borrowers and away from less well-connected borrowers, even when there is no difference in their productivity. Second, it makes it important that lending be local—the lender must know and trust his borrowers. This adds one or more layers of intermediation to the process of lending, with additional transaction costs entering at each of the stages, which raises the opportunity cost of capital. Third, it forces the lender to limit his lending, with the consequence that both his capital and his skills as a lender may remain unused for a significant part of the time. This raises both the opportunity cost of the capital and the transaction cost (which includes a part of the lender's time). Finally, it gives the lender some *ex post* monopoly power, as a borrower would find it hard to leave a lender who knows him well. Under competitive conditions, these *ex post* rents will be dissipated in *ex ante* competition, with lenders in effect subsidizing new borrowers in order to extract rents later from those who will become his long-term clients.

What this tells us is that the four components of the interest rate are all jointly determined in the process of the lender making his lending decisions. Depending on the lender's strategy, it could be that the transaction costs dominate or the opportunity cost dominates or that default or monopoly rents become very important. The strategy could be very different depending on the nature of the clientele and other environmental characteristics. This may be a part of the reason why different people have taken very different views of informal credit markets: Aleem, for example, finds that for every rupee lent, about half a rupee goes into transaction costs, while in Dasgupta (1989) only about 30% of interest costs are explained by transaction costs (strictly establishment costs) while Ghate (1992) argues that

[19] See McMillan and Woodruff (1999).

transaction costs are unimportant except in the case of very small loans.[20]

The fact that all these decisions are interrelated clearly makes it dangerous to use any single one of these components as a measure of the efficiency of intermediation. For example, Ghate (1992) sees the low level of transaction costs in his sample as evidence for the remarkable efficiency of informal lending. But, as has already been noted, transaction costs may be low because the lenders are very choosy about to whom they lend. This raises the opportunity cost of capital (since capital is often idle) and limits credit access, both of which have their welfare costs. Likewise, the low rate of defaults in informal transaction is often mentioned as evidence for their efficiency, but this is obviously misleading if it comes at the cost of increased monitoring or reduced lending. Finally, the presence of rents in lending is not, per se, evidence for lack of competition in the market. As pointed out above, in this type of market, *ex post* rents are consistent with *ex ante* competition.

A further implication of this observation is that both loan size and the interest rate are jointly determined and therefore one cannot give a causal interpretation of the relation between interest rates and loan size reported above. Rather, one should see both of these as outcomes that are determined by more primitive variables such as the wealth of the borrower, his productivity, the liquidity of his assets, etc. This also makes it harder to interpret the reported negative relation between the borrower's wealth and the interest rate. In principle, it could be entirely a result of the fact that rich borrowers borrow more.

Most importantly, this line of argument underscores the importance of developing a proper theory of credit markets. Such a theory would explain the variation in interest rates and the gap between interest rates and deposit rates in terms of the true primitives of the model, and make predictions about the relation between loan size and interest rates and borrower and lender characteristics. While there is a long tradition of models of imperfect credit markets going back to Jaffee and Russel (1976) and Stiglitz and Weiss (1981), and the arguments behind why credit markets can fail, based on moral hazard and/or adverse selection, are well-known, I feel that it is useful to develop a framework that has a more direct empirical orientation.

[20] Even within the set of case-studies reported by Ghate, there seems to be considerable variation. In Kerala, the case-study concludes that transaction costs are of negligible importance while the Thai study concludes that transaction costs added between 3 and14 percentage points to the interest cost.

3.2 A Simple Model of Moral Hazard in the Credit Market

There is an investment opportunity whose gross returns are $F(K)R(p)$ with probability p and 0 otherwise, where K is the amount invested and $F(\cdot)$ is a production function. If an investor wants to invest more than his wealth, W, he will need to borrow. There is a capital market and the (gross) cost of capital in that market is ρ. To make this problem interesting it is assumed that:

1. p is a choice for the investor but is unobserved by the lender. p takes a value between p_0 and p_1.
2. $E(p) \equiv pR(p)$ has the property that $E'(p_0) > 0$, and $E''(p) \leq 0$.
3. The only possible contract is a loan contract.[21]

3.2.1 The Basic Moral Hazard Problem

The optimal value of p, p^* is clearly greater than p_0 and may or may not be less than p_1. The combination of the rest of the assumptions tell us that there is no guarantee that p^* would be chosen in equilibrium. To see this, note that the borrower, who is assumed to be risk-neutral, will choose p to maximize $F(K)E(p) - pr(K - W)$, where r is the interest rate that has to be paid to the lender to make him willing to lend.

The borrower will choose p such that $E'(p)F(K) - r(K - W) = 0$.[22] This is quite obviously inconsistent with the social optimum: the borrower clearly wants to choose $p < p^*$. This is the standard incentive problem in credit markets: Society cares about net output but the borrower only cares about what remains after paying interest. This is the essence of all models of *ex ante* moral hazard in the credit market.

Next, notice that the first order condition for the borrower's choice of p can be rewritten in the form:

$$E'(p)\frac{F(K)}{K} = r(1 - \frac{W}{K}). \tag{1}$$

From this equation it is evident that p depends on three things: The average product of capital, $F(K)/K$; the leverage ratio K/W; and the interest rate, r. If capital is more productive, the borrower is less inclined to

[21] This rules out making the borrower's payments depend on the project's realized returns. Diamond (1989) justifies this assumption by assuming that the realized return is not publicly observable except by making use of a liquidation proceeding, which is costly to the point of using up all available output. This makes sure that a borrower will not willfully default as long as the lender threatens to go into liquidation whenever he defaults.

[22] Assuming an interior optimum exists.

misbehave, and this is reflected in a lower p. Being more leveraged worsens his incentives and so does a higher interest rate, which is consistent with the observation made above that the interest cost burden is the source of the distortion.

> **Property 1. Efficiency: There is less inefficiency in the credit relationship when there is less leveraging, when the interest rate is lower, and when the project is more productive.**

From this it follows that the equilibrium value of p can be written in the form:

$$p = p(R, F(K)/K),$$

where $R = r(1 - \frac{W}{K})$ is the interest cost per unit of investment. Clearly $\frac{\partial p}{\partial R} < 0$ and $\frac{\partial p}{\partial F(K)/K} > 0$. Writing the relation in this form draws attention to the important role played by the shape of the production function. When $F(\cdot)$ is concave, $F(K)/K$ decreases as a function of K. Therefore, those who invest more will be more liable to moral hazard, even after controlling for the leverage ratio. However, if F is convex, at least over a range, increasing the level of investment may increase profitability and improve the borrower's incentives. As we will see, this distinction may be very important for some questions.

3.2.2 The Interest Rate

We have so far treated r as a parameter. In fact, if there is competition in lending, lenders should not make any profits, which would imply that

$$r = \rho/p, \tag{2}$$

or

$$R = \rho(1 - W/K)/p = \frac{\Gamma}{p},$$

where Γ is the cost of capital per unit of investment.[23] Solving $p = p(R, F(K)/K)$ along with $R = \Gamma/p$, gives us $p = \widetilde{p}(\Gamma, F(K)/K)$ and $R = \widetilde{R}(\Gamma, F(K)/K)$.

[23] The assumption of perfect competition in the credit market is not uncontroversial. There is a long tradition of papers that view high interest rates as evidence for monopoly power in the credit market. However,as already pointed out, the issue of rents in the credit market is likely to be quite delicate, since competition operates *ex ante* rather than *ex post*. Therefore, the absence of *ex ante* rents is consistent with Bhaduri's (1977) model of how lenders trap borrowers into a permanent cycle of debt and debt repayment. The

However, it is easy to construct examples where these equations have multiple solutions: Intuitively, a fall in p raises r, but a rise in r, as we already saw, puts downward pressure on p. It is not clear, however, that we can interpreted these as multiple equilibria—if the lender knows the rules of the game, he knows that he can pick the best equilibrium and make everyone better off, simply by setting the right interest rate. Therefore, unless the lender is boundedly rational, we should probably assume that the best equilibrium is always chosen. This is the equilibrium with the lowest interest rate.

Assuming that this is the equilibrium, the comparative statics of the $p(\cdot)$ function are inherited by the $\widetilde{p}(\cdot)$ function, and $\widetilde{r} = \frac{\rho}{\widetilde{p}}$ shares the properties of the $\widetilde{p}$ function, only reversed. A lower leverage ratio increases p and lowers the interest rate, as does a higher average product of capital. Lowering the cost of capital lowers the rate of interest more than proportionately since the repayment rate goes up.

> **Property 2. Interest Rates: Borrowers who are more leveraged tend to pay higher rates, while more productive borrowers pay lower rates. Raising the cost of capital raises the interest rate more than proportionately.**

3.2.3 The Level of Investment

The next step is to endogenize the level of borrowing. The borrower's choice of K maximizes

$$F(K)E(p) - \rho(K - W)$$

under the assumption that p depends on K through the $\widetilde{p}(\cdot)$ function. The first order condition for that maximization is**:**

$$F'(K)E(p) + F(K)E'(p)\frac{\partial p}{\partial F(K)/K}\frac{\partial F(K)/K}{\partial K} + F(K)E'(p)\frac{\partial p}{\partial \Gamma}\frac{\rho}{W} = \rho. \quad (3)$$

If we compare this with the first order condition in a first best world, $F'(K)E(p^*) = \rho$, we see that there are three sources of distortion. First, $E(p) < E(p^*)$, which says that capital is less productive and therefore the borrower wants to invest less. Second, $\frac{\partial p}{\partial \Gamma}$ is negative which also discourages

evidence seems to support the hypothesis of *ex ante* competition: The few studies (Ghate (1992), Dasgupta (1989), Aleem (1989) that compute the gap between the interest rate charged and the various costs of lending (opportunity cost, monitoring costs, default costs) do not find a large gap on average, though one cannot reject the possibility that there is a large rent component in many individual transactions.

investment. Finally, there is the second term on the left-hand-side, which can be positive or negative depending on the sign of $\frac{\partial F(K)/K}{\partial K}$. This, as we have already observed, depends on whether the production function is concave or not. If it is concave, the second term is negative and it is unambiguously true that imperfections in the capital market lead to less investment. If not, the second term may be positive and if this effect is large enough, it could outweigh the first effect and generate over-investment. Whether this possibility is actually worth taking seriously remains an open question, awaiting more precise calibrations of the model.[24]

Another important property of the first best is that the amount invested is independent of the wealth of the investor. In our present model, if we were to increase W, keeping K fixed, we know from Property 1 that p would go up, raising $E(p)$ and reducing $E'(p)$. As long as F is concave, both of these effects go in the same direction: They both raise the rewards for investing more, and therefore there is more investment.[25] In fact, in the special case where $F(K) = \sigma K$, i.e., a linear production technology, K not only goes up when W goes up, it is precisely proportional to W.

The general case of a non-concave F tends to be complex. One interesting example, where there is a single indivisible investment, turns out to be very straightforward. In this case people either invest or do not, and since those who have more wealth choose a higher p at the same level of investment (Property 1), they are the ones who will invest. More generally, non-convex production technologies raise the possibility that the poor will actually invest more than the rich: Intuitively, if the production function is convex, increasing investment raises productivity, which improves incentives through its direct effect. However, there is also an indirect effect: Investing more makes the borrower more leveraged and this worsens incentives. The balance of these two effects may be different for the rich and the poor, since their incentive problems are different, and in principle it could be that the poor end up investing more. However, it seems unlikely that these effects would dominate the main effect of being richer, which is that (at the same level of investment) richer people are less leveraged and therefore have better incentives and as a result their capital is more productive.

Lowering the cost of capital in this model increases p and this, as argued above, encourages investment. However, lowering the cost of capital also increases the amount invested in the first best, so that there is no clear

[24] Lehnert, Ligon and Townsend (1999) argue that this is a real possibility.

[25] Actually, there is a third effect: Increasing W/K, it can be shown, reduces $\partial p/\partial F(K)/K$, thereby reinforcing the effect of the fall in $E'(p)$.

prediction for the extent of underinvestment.

> **Property 3. The Level of Investment: Capital market imperfections lead to under-investment in the typical case, though it is not inconceivable that they could generate over-investment. The more wealthy will tend to invest more in absolute terms. When the production technology is linear, the amount invested will be proportional to the investor's wealth. When there is a single indivisible investment, the rich are more likely to invest than the poor. Lowering the cost of capital increases investment.**

Capital market imperfections reduce the demand for capital. For a given supply curve of capital, this means that the cost of capital will be lower than it would be otherwise. In the longer run, however, the supply of capital will also respond to the pattern of wealth creation generated by the capital market imperfection and the net impact on the cost of capital is ambiguous.

> **Property 4: The Cost of Capital: For a given supply curve for capital, imperfect capital markets will have a lower cost of capital, but this is no longer necessarily true once we take into account the impact of the capital market imperfection on the supply of credit.**

3.2.4 Introducing Monitoring

The model developed so far is useful in developing intuition about how the credit market works but it has an important limitation in terms of explaining the data. As we have already seen, the repayment rates in most informal credit transactions are very high (over 90%). It follows from equation 2 that the interest charged by a competitive lender can only be about 10% higher than the cost of capital, which from all the evidence above is much too small a margin.

The missing piece of the story is monitoring. We have assumed so far that the lender cannot do anything to affect the borrower's choice of p. This is clearly an extreme assumption, since, as already mentioned, lenders can and do monitor borrowers.

The point of all these activities is to learn more about the borrower. This helps in two ways: First, by allowing the lenders to pick borrowers for whom the interval $[p_0, p_1]$ is relatively small, thereby limiting the possibility

of moral hazard. And second by getting to know the borrower's environment, thereby making it easier to find out when the borrower is not doing what he has promised to do with the money.

In addition to this kind of *ex ante* monitoring there is *ex post* monitoring of the project, checking that the borrower has done what he had promised to do with the money. For example, the lender can try to make sure that the borrower is spending the money on inputs for his project rather on consumption. Finally there is collection: Once the loan falls due, the lender has to spend time chasing each overdue loan.

It is not possible to capture all of these different aspects of monitoring in a single model, so we limit ourselves to one specific model, though some of the other models will be discussed in a later section. We introduce monitoring into the model by assuming that if the lender monitors at a level a, the borrower will chose a project $p(a)$ or a project with a p no lower than $p(a)$.[26] We assume that this comes about either through *ex ante* monitoring of the project (screening of projects before the loan is given) or *ex post* monitoring of the project (checking on the borrower after he has been given the loan and punishing him if he has not done what he was supposed to do). The problem is that we know very little about the nature of the empirical relation between monitoring and project choice. The only option we have is to reason on purely *a priori* grounds.

One assumption that has a certain plausibility is that the amount of monitoring necessary is a function of the extent of misalignment of incentives between the borrower and the lender. The borrower in our model wants to choose $p = p(R, F(K)/K)$, which gives him a payoff of $F(K)E(p(R, F(K)/K)) - p(K/W, r, F(K)/K)r(K-W)$, while the lender wants him to choose p which gives him a payoff of $F(K)E(p) - pr(K - W)$. The extent of misalignment is therefore

$$D = F(K)[E(p(R, F(K)/K)) - E(p)] - [p(R, F(K)/K) - p]RK. \qquad (4)$$

Our assumption is then that the amount of monitoring is a function of D. However, in order to allow for different types of scale effects, we write it in a slightly more general form:

$$M = M(K, D/K, m),$$

where m is a parameter that shifts the monitoring cost function ($\frac{dM}{dm} > 0$).

[26] The borrower will typically choose the lowest permissible value.

3.2.5 The Cost of Capital with Monitoring

The lender's participation constraint (2) now takes the form:

$$R = \frac{\Gamma}{p} + \frac{M(K, D/K, m)}{Kp}. \tag{5}$$

This equation defines $R(\Gamma, K, p, m)$, the interest rate for a borrower with a fixed W who wants to invest an amount K and promises to choose a project p. Using this, we can define the expected cost of credit per unit of investment: $C(\Gamma, K, p, m) = pR$.

This formulation of the supply side of credit has the obvious advantage that the interest rate can be much higher than the cost of capital even if defaults are rare. This is because monitoring costs can be very high—indeed the reason why there is very little default may be a result of the resources spent on monitoring.

It is useful to begin our analysis of this model with an examination of the properties of the $C(\cdot)$ function. Simple differentiation tells us:

$$\begin{aligned} \frac{\partial R}{\partial \Gamma} &= \frac{1}{p - \frac{p - p(R,F(K)/K)}{K} \frac{\partial M}{\partial (D/K)}} \text{ and} \\ \frac{\partial R}{\partial m} &= \frac{\partial M / \partial m}{K[p - \frac{p - p(R,F(K)/K)}{K} \frac{\partial M}{\partial (D/K)}]}. \end{aligned}$$

Since $[p - \frac{p - p(R,F(K)/K)}{K} \frac{\partial M}{\partial (D/K)}] < p$, this tells us that increases in the cost of lending (represented by a rise in ρ or in m) have a multiplier effect, resulting in a bigger increase in the interest rate than would be warranted by the direct effect of the increase in cost.[27] This is because the initial rise in the interest rate worsens the borrower's incentives and makes it necessary that he be monitored more, which raises the cost of lending even further, etc. This property is obviously also inherited by the $C(\cdot)$ function.

Property 5. Multiplier: The interest rate and the amount of monitoring can be very sensitive to changes in the cost of capital and/or the cost of monitoring.

This is an important property: It tells us that there are even relatively small differences in the monitoring cost, or the cost of capital can induce a lot of variation in the interest rate, which helps to explain why we observe so much variation.

[27] In principle, this increase can be very large since $[1 - \frac{p - p(R,F(K)/K)}{K} \frac{\partial M}{\partial (D/K)}]$ can be very close to 0 or even negative (in which case, the equilibrium interest changes discontinuously).

A related and important property of the $C(\cdot)$ function comes from differentiating equation 5 with respect to p. This, after some algebraic manipulations, gives us:

$$\frac{\partial C}{\partial p} = \frac{1}{K^2}\frac{\partial M}{\partial (D/K)}\frac{p(R, F(K)/K)RK - pF(K)E'(p)}{p - \frac{(p - p(R,F(K)/K))}{K}\frac{\partial M}{\partial (D/K)}}.$$

Equation 1 tells us that $E'(p(R, F(K)/K))F(K) = RK$. Using this (assuming that $p - \frac{1}{K}\frac{\partial M}{\partial (D/K)}(p - p(R, F(K)/K)) > 0$), it is immediately clear that the sign of $\frac{\partial C}{\partial p}$ depends on the sign of $p(R, F(K)/K)E'(p(R, F(K)/K)) - pE'(p)$. Since $p > p(R, F(K)/K)$ it follows that C_p can only be positive if the function $pE'(p)$ is a decreasing function of p over a range.

This makes it clear that it is entirely possible that C_p be negative for all p, implying that implementing high values of p may, paradoxically, require less monitoring than implementing lower values. This is because a high p generates a low R and this improves incentives.

In such situations it may be optimal to raise p all the way to its maximum, i.e. to p_1. In particular, this will be true as long as $E(p)$ is everywhere increasing in p over its admissible range[28] and it will remain true *irrespective of how costly it is to monitor*.[29] However, it is easy to see that it will never be optimal for p to exceed its social welfare maximizing level, i.e the value of p for which $E'(p) = 0$. This is because when $E'(p) = 0$, C_p is clearly positive.

Property 6. Default: Very low levels of default may be optimal even when monitoring is quite costly, though it is never optimal to have less default than in the social optimum.

This is important because it tells us that it is often optimal to aim for very low default rates even at the cost of lots of costly monitoring and high interest rates. This is reassuring, given that the combination of very low default rates and very high interest rates is by no means uncommon.[30]

3.2.6 The Optimal Credit Contract with Monitoring

The optimal credit contract will be a combination (K, p) that maximizes

$$F(K)E(p) - pR(\Gamma, K, p, m)KM.$$

[28] For example, C_p is negative whenever $E(p)$ takes the form Ap^β, with $A > 0$ and $\beta \in (0, 1)$.

[29] Of course, this is conditional on the loan contract being viable, which is not the case when monitoring is too costly.

[30] Aleem's data set from Pakistan, mentioned above, is an example.

The first order conditions that describe the optimal contract (when it is not a corner solution) are

$$F(K)E'(p) = KC_p \text{ and} \tag{6}$$

$$F'(K)E(p) = C + KC_K. \tag{7}$$

There is, however, relatively little that we can say about the optimal credit contract at this level of generality. The problem is easily seen from equation 5: An increase in K affects both the numerator and the denominator of the expression $\frac{M(K,D/K,m)}{(K-W)p}$, and without more structure it is not possible to say anything about how more investment affects the expected cost of lending.

The Model with Constant Returns in Monitoring One simple and fruitful way to impose structure is to assume constant returns in monitoring, i.e.,

$$M(K, D/K, m) = KM(D/K, m).$$

For the most part, we will also assume that there are constant returns in production, i.e., $F(K) = \sigma K$. In this case, equation 5 can be rewritten in the form:

$$pR = \Gamma + M(\sigma[E(p(R,\sigma) - E(p)] - [p(R,\sigma) - p]R, m).$$

R is therefore a function of σ, m, p and Γ, and so is, therefore, the expected cost of lending C. It follows that keeping p and Γ fixed, doubling the borrower's wealth and the amount he invests does not change the unit cost of lending. It follows that all borrowers with the same σ and the same m, will choose the same leverage ratio and face the same interest rate. In other words, under full constant returns the rich and the poor will pay the same rate of interest as long as they are equally productive. The rich will simply invest more.

The direct prediction of this model is the absence of a correlation between the borrower's wealth and the interest rate. Of course, as we will see later, it does not rule out the possibility of a spurious correlation, induced by a correlation between W and either σ or m. Nevertheless, this result provides a useful benchmark: It tells us that there is no necessary reason why the rich should pay lower interest rates, as observed in the data.

The Model with a Fixed Cost of Monitoring A model that manages to account for most of the observed fairly economically is one where there is a fixed cost of monitoring which has to be paid as long as there is some borrowing and a variable cost which, as before, exhibits constant returns i.e.,

$$M(K, D/K, m) = KM(D/K, m) + \Phi.$$

In this case, equation 5 can be rewritten to read

$$pR = \Gamma + \Phi/K + M(\sigma[E(p(R,\sigma) - E(p)] - [p(R,\sigma) - p]R, m). \quad (8)$$

With these assumptions, we run the risk that the lender's maximization problem may not be convex: To see why, note that $\Gamma + \Phi/K = \rho(1 - W/K) + \Phi/K$, which tells us that if $\Phi > \rho W$, R goes down when K goes up, encouraging the borrower to borrow even more. Conversely, a borrower who borrows little will pay very high rates, making it attractive for him to not borrow at all, suggesting the possibility of a "bang-bang" solution. As long as we make sure that σ is not too large (to avoid the possibility that the demand for credit becomes infinite), the solution will be for the poor borrower ($\rho W << \Phi$) to borrow nothing.

However, there is an interior solution for borrowers who are richer ($\rho W >> \Phi$). It is easily checked that this interior solution has a very simple property: From equation 8 it follows that as long as p, σ, ρ and m are held fixed, R is completely determined by the term $(\rho W - \Phi)/K$ and therefore $C \equiv C(\frac{\rho W - \Phi}{K})$. From 7 the optimal choice of K satisfies

$$\sigma = C(\frac{\rho W - \Phi}{K}) + \frac{\rho W - \Phi}{K} C'(\frac{\rho W - \Phi}{K})$$

in this case, which tells us that $\frac{\rho W - \Phi}{K}$ is uniquely determined by σ. A number of properties follow immediately from this observation. First, an increase in W results in a more than proportional increase in K: In other words, richer people are more leveraged. Second, higher values of Φ are associated with a lower value of K. Third, changes in W and Φ do not affect R, from which it follows that r goes down when W goes up (since K/W goes up and R remains unchanged).

There is a straightforward intuition behind these results. Given that there is a fixed cost of lending, those who invest more will face a lower cost of capital. However, for a poor person to be able to invest the same amount as a richer person, the leverage ratio would have to be much higher and this distorts incentives. The optimal contract balances these two types of costs: The poor end up both investing less and paying more in interest.

To assess whether the model generates the right orders of magnitude, the model was simulated under the assumption that $E(p) = 2p^{0.5}$, and $M(D/K, m) = mD/K$.[31] For parameter values $\rho = 1.05$, $m = 0.8$, $\Phi = 0.5$, $\sigma = 0.66$, we find that those with wealth levels up to about 1.75 (i.e., about three and a half times the fixed cost of monitoring) do not invest at all. When investment begins, the interest rate is above 50% (and the leverage ratio is 2.8) and as the borrower's wealth goes up, the interest rate goes down and converges to about 27%, while the leverage ratio rises and converges to about 4.2.

Property 7. Wealth Effects: When there are constant returns in both production and monitoring, two borrowers who differ only in their wealth levels will be equally leveraged and will pay the same interest rate. When there is a fixed cost of monitoring, richer borrowers will pay a lower rate and will be more leveraged and the very poorest borrower's will prefer not to borrow at all.

These wealth effects have the implication that the wealth advantage the rich start with will tend to get amplified: First, because the difference in the amount invested is typically going to be larger than the difference in wealth. This follows from the fact that the leverage ratio is greater than 1 and either constant or increasing in wealth.[32] Second, as long as R is increasing in K, the marginal product of capital will be higher than the expected interest cost in equilibrium and each unit of investment generates some pure profits (this follows from equation 7). Since the rich invest more than the poor, they earn more pure profits.

3.3 Taking Stock: How to Think about Credit markets

The model with a fixed cost of monitoring gives us a simple way of accounting for the facts that are listed above. Moreover, it helps to explain the fact that short run interest rates in informal markets are often higher than longer run interest rates (for example, see Table 3.2 in Dasgupta (1989)). Ghate (1992) also notes that very short short-term loans are often particularly expensive. In the face of it, this is puzzling because one would imagine that the scope for moral hazard is greater in longer term contracts. The fixed cost approach can resolve this puzzle as long as there is a part of the fixed cost that is

[31] Note that $pE'(p)$ is increasing in p so that it is always optimal to choose the highest possible value of p. We set this value to be 0.9, so that the default rate is fixed at 10% (which is high but within the observed range).

[32] Note however that this property may not hold if all the wealth was not liquid. In that case, the leverage ratio may be less than 1.

transaction-specific and independent of the length of the contract.

However, not surprisingly, this is not the only way to account for these facts. For example, if the production function were concave rather than linear, the rich would be less leveraged than the poor (since diminishing returns sets in at high levels of investment). As a result, the interest rate they face will tend to be lower. While they are less leveraged, the absolute amount that the rich borrow may still be higher. High levels of credit will therefore be associated with lower interest rates. Similar patterns may arise if, for example, it is cheaper to monitor the rich because, say, the rich share closer social ties with those who are lending.

In any case, there is no reason why we should confine ourselves to models of *ex ante* moral hazard. *Ex post* moral hazard (borrowers who try to avoid repaying) is clearly an important aspect of credit markets and so is adverse selection.[33] There are also other ways to model *ex ante* moral hazard: Holmstrom-Tirole (1996) develop a model where the borrower wants to put in less effort rather than take too much risk. Many of the basic predictions of our model show up in these models as well. Not surprisingly, less leveraged borrowers, all else being the same, tend to be better credit risks and face lower interest rates at the same level of borrowing. Consequently, richer borrowers will have their initial advantage compounded by the workings of the credit market. High interest rates, as before, make the borrower more likely to misbehave in these models as well.[34] This, in turn, tends to raise the interest rate (either because default becomes more likely or because more monitoring is called for). Therefore, what I call the multiplier property of interest rates—namely the fact that small increases in the costs of lending can lead to a large increase in the interest rate—ought to be true in these other models as well. Finally, while the cost of monitoring is rarely formally introduced into these models (Holmstrom-Tirole, 1996, being an important exception), it is intuitively clear that the shape of the monitoring cost function will play a crucial role and a fixed cost of monitoring will have an effect similar to the one discussed above.

One might even want to venture beyond models where the borrower is the main source of moral hazard. The paper by Holmstrom and Tirole, mentioned above, worries about the incentives of intermediaries in the credit

[33] Hart-Moore (1994) is the best-known model of *ex post* moral hazard in the credit market, while Stiglitz and Weiss (1981) is the classic reference on adverse selection.

[34] Though as pointed out in Aghion-Banerjee-Piketty (1999), in models of *ex post* moral hazard, there is a possible countervailing effect coming from the fact that high interest rates make it more credible for the lender to put a lot of effort into pursuing recalcitrant borrowers.

market. They argue that these problems translate into a further credit constraint, this time at the level of the intermediary. Moreover, the typical intermediary is large, and if the intermediary itself has the right incentives, its agents who do the actual business of lending–the loan officers and credit appraisers of the world–may not, since what they have at stake personally is only a very small part of the amount of money they control. The solution is typically to restrict the domain they control: Stringent bureaucrat rules about what loan officers can and cannot do, are a feature of credit market intermediaries the world over. This, of course, comes at a cost. Credit decisions become bureaucratized and typically much less responsive to all but the hardest information. Assessments of the quality of the project and judgements about future profitability, both relatively soft information, will tend to have little impact on who gets credit. To the extent that other institutions, such as venture capitalists, do not pick up the slack, this will hurt new entrants and the most radical ideas. This problem may also be most serious where the banking sector is dominated by the public sector, given that there is already a tendency towards bureaucratization and buck-passing.[35]

Of course much more work (simulations, etc.) is called for before we can be sure that these alternative theories can generate the right orders of magnitude. But in one sense there is no reason to pose these as alternatives. All of them working together can generate a larger aggregate effect, and large aggregate effects are clearly important in giving relevance to this class of theories. However, from the point of view of actually designing policy it is important to know exactly where the constraint lies. It is also, both from the point of view of macro-relevance and micro-policy design, important to identify the exact structure of the credit constraint. Is the amount of credit that a borrower can get primarily a function of the borrower's wealth and his expected profitability, as our model suggests, or is it the case that profitability is largely ignored, as the simple bureaucratic model sketched above would suggest? How important is the distinction, ignored in our analysis between wealth and liquid wealth—one additional reason why the poor may suffer is that their wealth may be less liquid?[36] Is the borrower's

[35] For evidence that this is a very real problem, see Banerjee and Duflo (2001). There are also lots of anecdotes that support this view: It is said, for example, that Indian bankers in the 1980s and early 1990s were puzzled by how they could justify lending to software companies, since their only real assets were their staff and work-in-progess was lines of code on the computer.

[36] The work by Swaminathan, already cited, suggests that this may indeed be an important distinction. The requirement that the wealth be held as liquid collateral also

wealth the key ingredient or is it his inventories, as in the classic models of inventory financing?[37] The answer to each of these questions has many important ramifications, and careful empirical research on the technology of lending remains one of the imperatives of the day.

Theme 1: The observed patterns in credit markets—low default rates, high and highly variable interest rates and credit limits that increase with wealth—suggest that contracting in credit markets is highly imperfect and monitoring is very important. This suggests that there will be under-investment, and a significant part of the output produced will be wasted on monitoring. The earnings gap between the rich and the poor will be amplified by the capital market imperfection and this will be particularly the case if, as seems plausible, there are fixed costs of monitoring.

4 The Dynamics of Wealth Accumulation

One important message from the previous section is that the poor are at a disadvantage in the credit market. However, as we have seen, the exact form of the disadvantage tends to depend on the technologies of production and monitoring. For example, in the case where there are constant returns in both production and monitoring, the disadvantage takes the form of a proportional reduction in the amount they invest with no difference in the interest rate or the choice of projects, whereas a model with a fixed cost of monitoring generates variations in the interest rate and in the choice of projects. In the short run, however, all versions of the capital market imperfection have the common implication that the poor will be underrewarded for their talent.

The longer run implications of the different models are, however, potentially very different. To see this, imagine a world where there is one good produced and a population of identical people who each live for one period and always have one child. Each person starts life with an endowment that her parent gave her. Her life is simple, verging on the drab. At the beginning of her life, she chooses among income earning opportunities. Her choices will be to either invest in a productive opportunity or to lend out the money.

creates a demand for collateralizable assets and shifts in the relative price of these assets become important (see Kiyotaki and Moore, 1997, for a macroeconomic model based on this particular relative price shift).

[37] Based presumably on the idea that loans have to be fully collateralized and inventories can serve as collateral. Clearly the plausibility of this model depends on how easy it is to attach inventories.

The exact technology of production will be discussed later, but it could be thought of either as investing in learning a skill, starting a business or even patenting or promoting a new idea. At the end of the period, she decides on what to do with her realized income, which consists of her investment earnings plus an endowment e – she can leave it to her child or eat it herself. For simplicity, assume that she has Cobb-Douglas-like preferences over consumption (c) and bequest (b):

$$U(c,b) = A[c^{1-\beta}b^{\beta}], 0 < \beta < 1, A > 0.$$

Since she allocates her end-of-period wealth between these two uses, this immediately implies that if her end-of-period income (or wealth) is y:

$$c = (1-\beta)y, \qquad b = \beta y.$$

People borrow to invest more than what they were born with. Assume that the credit market is exactly as modeled in the previous section.

These assumptions together define a simple dynamic process that maps W_t, the wealth of an individual from the current generation (which is equal to the bequest he received), into the wealth of his child, W_{t+1}. The exact shape of this map will depend on what we assume about technology, which is what we turn to now.

First consider the case where both the production technology and the monitoring technology are fully linear. In this case we know that the optimal leverage ratio, the optimal level of monitoring, and the optimal choice of p are all independent of the investor's wealth. To simplify life, assume that the optimal choice of p is 1. Denote the monitoring cost required to sustain this by m (per dollar lent). Finally, assume that the production technology takes the form $F(K) = E(1)K$ and that the optimal leverage ratio is λ. In this case, the map from current wealth to future wealth is given by

$$W_{t+1} = \beta[e + \kappa W_t],$$

where $\kappa = E(1)\lambda - (\lambda - 1)(\rho + m)$.

In the case where $\beta\kappa < 1$, this is a process that converges to a wealth level $\beta e/(1-\beta\kappa)$: Every dynasty ends up with the same wealth in the long run. On the other hand, in the case where $\beta\kappa \geq 1$ no dynasty ever converges but in the long run the wealth of every dynasty grows at the same rate and everyone becomes extremely wealthy. There are no poverty traps in this model.[38] The dynamics of wealth in this model are shown in Figure 1.

[38] However, this model is consistent with divergence across countries: Those countries that have a better financial system, and therefore a lower m, will grow faster.

Insert Figure 1 here.

The model behaves in much the same way when the production and monitoring technologies are strictly convex (i.e., both functions are strictly concave). A simple way to see this is to note that the poor always have the option of choosing the same leverage ratio as the rich. If they were to choose the same leverage ratio, they would actually pay less in interest than the rich, since they have a higher marginal product of capital and monitoring them is easier (diminishing returns in monitoring). Their net return per dollar of their wealth would therefore be higher and as a result, the wealth of the poor will grow faster than that of the rich.[39]

Things change significantly when at least one of the technologies stops being convex. To take an example, assume that the production technology remains what it was in the previous example but introduce a fixed cost of monitoring. Under these assumptions the equation representing the evolution of wealth for those who can borrow can be written in the form:

$$W_{t+1} = \beta[e + E(1)K_t - (K_t - W_t)r], = \beta e + \beta\lambda(W_t)(E(1) - R).$$

As shown in the previous section, when there is a fixed cost of monitoring R is constant but λ is an increasing function of wealth and converges to a $\overline{\lambda}$ as wealth continues to go up. For wealth below some W^*, people do not borrow and their wealth evolves according to

$$W_{t+1} = \beta e + \beta W_t E(1).$$

It is easy to see that this generates a map which has the convex shape in figure 2.

Insert Figure 2 here

The wealth dynamics implied by this picture are very different from those generated by the linear model. The rich get richer all the time but the poor converge to a steady state at $\widetilde{W}$. However, it is not neccessary that inequality grows without a bound: It is straightforward to add to this model

[39] This claim does depend in an important way on the fact that there is only one type of investment. Mookherjee and Ray (2000) show that if there are many alternative types of investment and the economy needs all of them, long run inequality may be inevitable in a world of imperfect capital markets, even if all the individual technologies are convex and people are forward-looking in their savings decisions.

a technological limit on investment, $\overline{K}$. It is easily checked that what this does is to reduce the amount borrowed by those who would have otherwise invested more than $\overline{K}$. Eventually, when someone's wealth exceeds $\overline{K}$, they will stop borrowing and start lending, which causes the W_{t+1} schedule to take the form

$$W_{t+1} = \beta[e + E(1)\overline{K} + (W_t - \overline{K})\rho].$$

As long as $\beta\rho < 1$, this implies that wealth will remain bounded: By the fact that the curve is continuous, it follows that it must be S-shaped. Depending on the parameters, this model has either one or three steady states, of which the two extreme ones are stable. When there are two stable steady states, the poor converge to the lower one and the rich to the upper, reflecting the fact, noted above, that when there is a fixed cost of monitoring the poor earn a lower return on their wealth.

Insert Figure 3 here.

This is the classic poverty trap situation: The poor remain poor and the rich stay rich. Versions of this story have been told in many papers, including Galor and Zeira (1993), Dasgupta and Ray (1986), and Banerjee and Newman (1994), though they each have a different name for the investment: education in Galor-Zeira; health in Dasgupta-Ray; and capital in Banerjee-Newman. Both Galor-Zeira and Dasgupta-Ray have explicit non-convexities in the production function while Banerjee-Newman introduces the non-convexity into the technology of lending. Moav (1999) presents a slightly different version of the argument which relies on non-convexities in what one might call the bequest technology.[40]

Of course there are good reasons not to take poverty traps literally. The very lucky and the very talented among the poor will probably manage to escape their background, and some of the rich will surely manage to squander their patrimony. The robust implication of this model is rather that economic mobility will be slow. An economy that starts with a lot of poor people will remain both poor and unequal for a very long time. This, however, raises the question of whether, in the end, this model is very different from the model where all the technologies are convex and there is no poverty trap, but the capital markets are extremely inefficient—after all

[40] In terms of our notation, the assumption he makes is that the amount bequeathed is a strictly convex function of end-of-period income over some range, whereas we have so far assumed that it is a linear function. It is easy to see that this can generate a poverty trap–essentially the rich benefit from the fact that rich parents leave bequests that are disproportionately large relative to their wealth.

when capital markets do not work very well, convergence can take a very long time.

My view is that it is nevertheless worth distinguishing between these two models because the forms of mobility that they permit are rather different. The convex model predicts a slow and steady rise for all the poor which culminates in their catching up with the rich (or if there is no convergence, they still become very, very rich). The mobility in the model with non-convexities, by contrast, comes from those who are either very talented or very lucky. In other words, it takes the form of large jumps by a relatively few people.[41] This difference also shows up in the shape of the long run distributions. In the convex model, most people in the long run will be "middle-class" with some outliers who are either very lucky or very unlucky. By contrast, in the other model most people will either be rich or poor.

Theme 2: Models where all the key technologies—the production technology, the monitoring technology and the bequest technology—are all linear or convex tend to favor long run convergence: Those of comparable talent will earn comparable amounts in the long run and the long run distribution of income will reflect the distribution of abilities in the population. By contrast, models where at least one of these technologies is non-convex can generate poverty traps at the level of the individual. People who start poor stay poor, and therefore equally able people may earn very different amounts even in the long run.

4.0.1 Endogenous Savings and Poverty Traps

The one obviously unsatisfactory aspect of our model so far is the modeling of bequest decisions. While the way we have modeled them is extremely convenient and there is no good reason, either empirical or *a priori*, to switch to full "Barro" preferences, there are clearly cases where our model seems a bit strained. In the specific example in the previous section where monitoring has a fixed cost, the rate of return on beginning-of-period wealth (i.e., bequests) varies enormously and those who start with more will, over a range, earn a higher return. It is therefore plausible that parents will take this into account when planning their bequests.

To see what changes when we allow for endogenous savings decisions,

[41] Paulson and Townsend (2000) and Jeong and Townsend (2000) make a related but different distinction between models (such as Evans and Jovanovic, 1989) where poor capital markets hurt the most talented people and models (such as Lloyd-Ellis and Bernhardt, 2000) where poor capital markets hurt most the least talented.

consider a modification of our basic model that makes the agents infinitely-lived and endows them with the standard forward-looking preferences. In the case where the production technology is convex and credit markets are absent, this is a special case of the model studied by Loury (1981) where he showed that there is convergence despite the credit markets being absent. Indeed, as emphasized by Caselli and Ventura (1996), the presence of effective credit markets in such an environment may actually slow or even stop convergence. The point is that poor capital markets can act as a spur to savings, since they make it more important to have one's own wealth and this effect is strongest for the poor, since capital is most productive in their hands. By contrast, when capital markets are perfect the marginal product of capital is equalized everywhere and the poor have no more incentive to save than the rich.[42]

When we combine this model of savings with a non-convex monitoring technology, things change dramatically. As noted above, the rate of return on savings is now lower for the poor than for those who are somewhat richer. This is especially true of the very poor, who cannot borrow at all. As a result, the rich (or at least the middle classes, since the very rich also earn low returns on their savings) will save a higher proportion of their income than the poor, which reinforces the poverty trap.[43]

Theme 3: If savings decisions are based on future benefits, capital market imperfections can actually promote savings and convergence. However, if the technology of production or monitoring is non-convex, the encouragement effect may only operate on the relatively wealthy. The poorest get low returns from capital and therefore will save relatively little, which may reinforce the poverty trap.

4.0.2 Endogenous Prices and Collective Poverty Traps

One implication of there being a poverty trap at the level of the individual is that there is also a collective poverty trap: an economy that starts with a lot of poverty will end up with a lot of poverty. *Collective poverty traps can, however, exist in models with imperfect credit markets, even when there are no individual-level poverty traps.* This point, first noted by Banerjee

[42] As pointed out by Ghatak, Morelli and Sjostrom (2000), it can also act as a spur to hard work, as a way to accumulate capital.

[43] This is strictly only true if we assume that the average savings rate for the economy is still β, but now the poor and the rich have different savings rates.

and Newman (1993),[44] relies on the fact that in a world where people are credit-constrained, factor prices depend on the wealth distribution (since the demand for factors depends on who has how much wealth). But the wealth distribution in any economy depends on factor prices—this two-way interaction creates the possibility of multiple steady states.

To see exactly how this might happen, consider a variant of the model developed above with linear production and monitoring technologies. The one factor price in that model was the cost of capital, which, so far, we took as given. To endogenize the interest rate, assume that the supply of capital comes from a fixed fraction of the population who cannot invest in the linear technology. Assume that in every generation, a fraction μ are handicapped in this way, but this attribute is neither correlated over time nor correlated with their wealth. These people (and everyone else) do have an alternative investment possibility which, for want of a better description, we will call "land": The return from investing in land is given by a strictly concave production function $H(\widetilde{\mu K})$, where $\widetilde{K}$ is the average amount invested in land by each investor who invests in land. Since this is a completely safe investment it will earn the safe rate, ρ, on the marginal unit, i.e., $\rho = H'(\mu\widetilde{K})$. This allows us to write $\rho = \rho(\widetilde{K})$, $\rho' < 0$. The rest of the available wealth in the economy will be invested in the linear production technology. In other words, if $\overline{W}$ is the per capita wealth in the economy, and if $\overline{K}$ denotes the average amount that each investor puts into the linear production technology, the market clearing condition for the credit market will be:

$$\mu\widetilde{K} = \overline{W} - (1-\mu)\overline{K} = \overline{W} - (1-\mu)\lambda(\rho(\widetilde{K}))\overline{W}, \tag{9}$$

where $\lambda(\rho)$ is the optimal leverage ratio from the point of view of borrowers in an economy where the cost of capital is ρ.

How about the evolution of wealth in this economy? Given all the assumption that have been already made, this turns out to be quite straightforward:

$$\overline{W}_{t+1} = \beta[e + H(\mu\widetilde{K}(\overline{W}_t)) + (\overline{W}_t - \mu\widetilde{K}(\overline{W}_t))(\sigma E(p(\rho(\widetilde{K}))) - M(\rho(\widetilde{K})))],$$

where $p(\rho)$ is the optimal choice of p when the cost of capital is ρ and $M(\rho)$ is the corresponding level of monitoring per unit of capital. From equation 9, when $\overline{W}$ goes up, $\widetilde{K}$ also goes up, but less than proportionally since ρ goes down and λ goes up. In other words, as the economy gets richer, a higher and higher proportion of its wealth will be invested in the

[44] There can also be collective poverty traps in models where there are no credit market imperfections, if, for example, there are peer group externalities (see Durlauf, 1996).

linear production technology. Since the net return to the linear production technology, $\sigma E(p(\rho)) - M(\rho)$, is greater than ρ (from equation 7) the average return on capital may go up as a result of the shift between the two sectors. This tells us that the $\overline{W}_{t+1}(\overline{W}_t)$ map need not be concave. In particular, it can have the S-shape depicted in Figure 4. This is most likely if it is the case that at low levels of $\overline{W}_t$ most of the capital is invested in land, but when $\overline{W}_t$ goes up beyond a certain point, the marginal product of capital invested in land falls off very quickly and, as a result, all additional capital is allocated to the alternative technology.

The S-shape in Figure 4 generates what one might call a collective poverty trap. No individual in this economy is ever trapped in poverty since all the technologies are convex. Nevertheless, economies that start poor stay poor. This is because capital is expensive in these economies and we know from above that when capital is expensive it is more likely to be used wastefully (our Property1).

This argument is an instance of a very general point: In all of these models, the distribution of wealth determines the pattern of investment, which in turn determines the demand for various factors, the factor prices, and eventually the next period's distribution of wealth. Since the effect of demand on factor prices is usually non-linear, the map from the present distribution of wealth to the future distribution so generated will typically also be highly non-linear, and therefore there is no reason to expect unique steady states in dynamic models with endogenous price determination.[45] Indeed, there is no presumption that these dynamic processes necessarily converge to a steady state: Aghion, Banerjee and Piketty (1999) and Aghion, Bacchetta and Banerjee (1999) generate endogenous cycles from models of this class.

This type of argument can also be made using other prices: Banerjee and Newman (1993), who were the first to emphasize this type of multiplicity, relied on the endogeneity of wages. High wage economies, in their argument, allow the children of the poor to become entrepreneurs, with the result that the demand for labor and wages remains high.

What is key in all of these cases is a perverse price effect: In the presence of capital market imperfections, price changes can have powerful wealth effects which make it possible that an increase in the price actually leads to an increase in the excess demand. Thus, in the just-mentioned paper by Banerjee and Newman an increase in the wage raises the demand for capital, while in the example represented in Figure 4, a rise in the interest rate raises

[45] For a more elaborate discussion of this point, see Banerjee and Newman (1993).

the excess demand for capital.

Perverse price effects are, however, no guarantee that global convergence will fail. Aghion and Bolton (1997), who were the first to analyze a dynamic model of exactly this class (i.e., with credit market imperfections and an endogenous interest rate), had focussed on the case where there was global convergence. Their point was that the process of convergence involved a Kuznets-like sequence of increasing and then decreasing inequality. Our example above is inspired by the work of Piketty (1997) who showed that the Aghion-Bolton model can, under suitable parametric assumptions, generate multiple steady states.[46] The key difference seems to come from assumptions about the elasticity of demand for capital: In Aghion-Bolton the demand for capital is relatively elastic and therefore there are no sharp changes in the interest rate. As a result, the economy is always quite stable.[47]

Theme 4: There can be collective poverty traps even when there are no individual poverty traps. This is most likely to be the case if the supply and demand for factors are not too price-elastic.

4.1 Taking Stock: How Plausible Are Poverty Traps?

There are two parts to the answer to this question. First, are the assumptions of the poverty trap model plausible? Second, do the implications of the model correspond to something we observe? These are addressed one by one.

One key assumption of the model of the poverty trap at the level of an individual is a non-convexity in either the production function, the monitoring function or the bequest function. Of the three, non-convexities in the bequest function are perhaps the easiest to document. Empirical evidence

[46] In Piketty's example (as in our's), multiple steady states arise because high interest rates are inefficient and this inefficiency reduces the supply of capital, and therefore the interest rate goes up. Interestingly, Matsuyama (2000) who provides an alternative argument for multiple steady states in this class of models, relies on an argument that associates high interest rates with *efficient* steady states. Raising the interest rate in his model allows the poor to accumulate wealth faster (since they tend to be lenders) and this increased wealth allows more poor people to transition faster into being investors, raising the demand for capital and the interest rate. In other words, while both Piketty and Matsuyama build their argument on the fact that in this type of model the excess demand for capital can go up when the interest rate goes up, in Matsuyama this happens because the demand for capital goes up while in Piketty it is the supply that goes down.

[47] While high interest rates are typically associated with underdevelopment—which is what Piketty's model tells us—neither of these models is intended to be "taken to the data".

from many OECD countries supports the view that bequests are a luxury good. Only the richer people leave bequests of any significant size so that the bequest function, at least over a range, is clearly convex.

Non-convexities in production are certainly very plausible. Most machines have a minimum efficient scale, and while rental markets can ameliorate this problem, they provide, at best, an imperfect substitute.[48] In the case where the good being produced is a usable education, there are several potential sources of non-convexities: Learning the letters of the alphabet is probably useful only when it translates into the ability to read simple sentences. For this reason, the first few years of education may not generate any returns, unless instruction is continued further. This is what Card and Krueger (1992) find in U.S. data: The first five years of education have no direct return. In the U.S., of course, very few people plan to get less than five years of education, but for some developing countries this may be an important non-convexity. Angrist and Krueger (1999), once again using U.S. data, find significant jumps in the return to education at school completion and college completion. Case and Deaton (1999) also report that the relation between wages and years of schooling for black South Africans starts relatively flat and becomes steeper. However Psacharopoulos (1994), concludes in favor of concavity, based on range of studies from all over the world and Duflo (2001) finds an essentially linear relationship based on data from Indonesia. Assesing the relative merits of these studies is beyond the scope of this paper but it is clear that the definitive paper on the important issue of non-convexity in education is yet to be written. Note however that in any case the fact that there a non-convexity is only interesting if it is large enough to be a hurdle for a significant number of people: Ideally, we would like evidence that allows us to scale the wealth distribution among potential investors in an industry to the size of the non-convexity in that industry. This would allow us to answer questions such as: What fraction of the population of potential investors in this industry are sufficiently poor that the non-convexity is relevant for them?

The situation in the case of the monitoring technology is even worse because very little is actually known. Aleem only reports the average monitoring cost for the entire population, which does not tell us how the cost changes with the amount borrowed. There are certainly *a priori* good reasons to suspect that some part of the cost—such as the cost of meeting the first time with the potential borrower—is a fixed cost. Moreover, as already discussed, fixed costs provide a natural explanation of the observed patterns

[48] Machines that require careful maintenance are typically not available for rent.

in credit markets. But we certainly need direct evidence about the size of the costs that are incurred by the lender at different stages of his relationship with borrowers and how that varies across borrowers.

There is also an *a priori* argument against non-convexities. Lehnert (1998) has argued that wherever there are non-convexities, people should participate in lotteries, which will make all of them better off and eliminate both the non-convexity and the poverty trap. This is not as fanciful as it might seem. Many poor people throughout the world participate in ROSCAs, which, it has been argued, are a type of private lottery designed to deal with non-convexities.[49] However, ROSCAs typically have the feature that all participants get to "win" over a relatively short period of time (at most a year) which is feasible when the non-convexity is of the order of magnitude of a year's savings for the average participant but not for larger non-convexities. This may reflect the fact that people are unwilling to lose a large amount of money in a single lottery, perhaps because there is loss-aversion or regret in their preferences. Moreover, the argument only applies to the non-convexities in production and monitoring. What is described as a non-convexity in the bequest technology is actually a simple non-linearity and is not subject to this criticism.

There is also the possibility that as the economy becomes richer and richer (through a process of technological upgrading, for example), the non-convexities will become less and less important. For this to be true, the level of the non-convexity must grow less fast than the income of the average person: This is unlikely to be true if, for example, the non-convex cost is a labor cost (the cost of monitoring, for example). Moreover, the process of technological upgrading is often accompanied by an increase in the size of the fixed cost.

Non-convexities, by themselves, do not guarantee that there is a poverty trap. What we need for an individual level poverty trap, to put it crudely, is evidence that the map from current wealth to future wealth is steep enough to cut the 45° line more than once. This can be a tough criterion to meet. For example, Dasgupta (1993) has argued, based on evidence from studies by biologists and nutritionists of the effects of malnutrition, that the relation between the parent's health (his proxy for wealth) and the child's health tends to be highly non-linear and typically includes a non-convex section. While this is important evidence, it is not enough to establish the existence of a poverty trap as interpreted here.[50] The problem is that the available

[49] See Besley, Coate and Loury (1993).

[50] Dasgupta (1997) suggests a more inclusive definition of a poverty trap in response to

estimates of the elasticity of income with respect to nutrition and health, as well as the elasticity of nutrition with respect to income tend to be less than 1,[51] suggesting that there cannot be an individual level poverty trap.[52] On the other hand, it can be argued that this is of course only one of many mechanisms that go into the map from current to future wealth. The question then is whether these mechanisms tend to reinforce each other or cancel each other out—note that they can cancel each other out even if there is no actual interaction between the various mechanisms, simply through a process of averaging.[53] This seems to be an important area for future research.

The other pillar of this class of theories are wealth effects on investment. The evidence on wealth effects on access to credit has already been discussed at some length and it may be presumed that these translate into wealth effects on investment. However, it is not clear that the causal factor here is necessarily wealth rather than some correlate of wealth. More direct evidence is now available from studies of firms in the U.S.,[54] showing that firms that get positive cash-flow shocks invest more, even after controlling for changes in their productive opportunities. In terms of developing country data, there is a recent study by Jeong and Townsend (2000), based on data from Thailand, which shows that the probability of owning a non-farm business is less than 10% in the bottom decile and over 30% in the top decile.[55] It not clear, however, that the entire effect they find is attributable to the direct effect of wealth rather than the effect of characteristics that are correlated with wealth. Banerjee and Duflo (2001) use a change in directed lending policies in India as a natural experiment to estimate the effect of greater access to working capital on profits. They find that an extra *one rupee of credit increases profits, net of interest, by more than one rupee.* Duflo (2000) finds evidence of strong income effects on investment in the health of young girls in South Africa, though it is not clear that it primarily

this kind of criticism.

[51] Strauss and Thomas (1993) list more than twenty studies that estimate the relation between income and calorie consumption (as a meaure of nutrition). Each one of them reports an estimate of less than 1 and most are less than 0.5. Strauss (1986) estimates the elasticity of wages with respect to nutrition in Sierra Leone and reports an elasticity of about 0.5.

[52] Srinivasan (1994) makes a similar point.

[53] For example, a series of non-convexities at slightly different places may average to a convex map.

[54] See, for example, Fazzari, Hubbard and Petersen (1988), and Fazzari and Petersen (1993).

[55] Evans and Jovanovic (1989), show a similar result for the U.S.

reflects the effect on access to credit. However these studies are relatively recent and only time can tell whether their results will turn out to be robust enough to be the foundations of a theory of development. As of now, as in many other instances, theory seems to be ahead of the evidence.

Collective poverty traps, unlike poverty traps at the level of the individual, do not directly rely on non-convexities. Wealth effects on investment are, however, clearly crucial as are strong price effects, which in turn rely on demand and supply for factors being relatively inelastic. While we do have some estimates of these elasticities (though mainly from developed countries), the theory is yet to be developed in a form that would allow calibration using these estimates.

The other approach to evidence would be to look directly for poverty traps. In other words, we could study the rates of economic mobility at different levels of wealth. Or we could look at the evolution of the wealth distribution and try to estimate from it the implied parameters of the underlying economy (including mobility rates). Robert Townsend, in joint work with a number of his co-authors, has done interesting work along these lines.[56] The problem remains however that we have no independent estimate of the part of mobility that arises from sources that are excluded from our model, such as genetics or learning.[57] For this reason, a more promising approach may be to look at the mobility patterns within a single industry, where it may be easier to calibrate some of the other sources of mobility.

There have been also been attempts to use cross-country data on inequality and growth to look at this question.[58] These exercises suffer from two problems. First, a number of theories, including those discussed here, predict a relation between inequality and growth, and the data does not distinguish between these alternative channels. Second, as argued by Banerjee and Duflo (1999), there is severe danger of misspecification, both because of possible omitted variables and strong non-linearities in the data.

4.2 Beyond Poverty Traps

Poverty traps are only the starkest form of what makes the world with imperfect credit markets interesting. As already noted, the more general phenomenon of slow convergence and limited social mobility is both interesting and important in itself. There are a number of other interesting predictions

[56] See Jeong and Townsend (2000).

[57] There are attempts to measure the contribution of genetics to economic mobility. See Bowles and Gintis (2001), for example.

[58] See Benabou (1996) for an excellent review of this literature.

of the imperfect credit market model, some of which we now briefly sketch.

First, poor capital markets will tend to be associated with great diversity in firm size within the same industry. In the case where there is an optimal scale of production, we would expect to see firms that are both below and above this scale—the former because they are capital starved and the latter because overall underinvestment generates rents in the industry and as a result those who do have the capital overinvest in order to capture those rents. This appears to be consistent with the oft-remarked fact that most industries in developing countries have both large firms and a fringe of very small firms.

Second, one might expect to find some very diversified firms in economies with poor capital markets. This because there will be rents in most industries and those who have money to invest can capture these rents. Therefore, the extent of diversification may not be guided by the usual considerations of competence and synergies. The Trading Companies in Japan, the Chaebols in Korea and the Managing Agencies in India are potential examples of this phenomenon.

Third, industry may be clustered in certain locations that have little to recommend them as venues for that industry. This is because informal lenders prefer to lend to borrowers they can monitor easily and this might lead them to prefer those who invest locally and in familiar industries. In Banerjee and Munshi (2001), we argue that the concentration of the knitted garment industry in Tirupur, an otherwise unknown town with poor infrastructure, is partly a result of the fact that the local population of Gounders need an outlet for their agricultural surplus.

Finally, tied transactions will be very important in this world, because tying saves on the costs of monitoring.[59] Thus, sellers will be the preferred source of credit for buyers (trade credit) and employers will be the preferred lender for workers. Cunat (2000) justifies the persistence of expensive trade credit in OECD countries (the standard rate in the U.S. is 44%) in these terms.

5 Other Contracting Problems

Contracting problems in the land market are, not surprisingly, very similar to the problems in credit markets. The models predict that wealthier tenants

[59] This argument goes back to Bardhan (1983).

will get more efficient contracts and more land.[60]

Contracting problems faced by lenders are, of course, just the other side of the contracting problems faced by borrowers. But the models of poverty traps typically focus on the borrowers even though it is easy to see how a very similar story would apply to lenders. In this story, the poor will not be able to lend to individual borrowers because there are fixed costs of monitoring, but intermediaries in the credit market also do not want to deal with them because there is a fixed cost of collecting money from them (they have to meet, for example). Therefore, they earn very low returns on their savings: Rutherford (1999) documents numerous examples where the poor accept substantial negative rates of return in order to put their savings in a safe place. Given the low return on savings, they prefer not to save and stay poor.

This story is empirically at least as plausible as any other story of the poverty trap, especially given that the poor are more likely to be lenders than borrowers. Yet there are few models of this type. Matsuyama (2000) is one paper that does take this issue seriously, but his focus is on collective rather than individual level poverty traps.

Contracting problems in the insurance market, by contrast, can be quite unlike those in credit markets. Insurance is a key market for the poor because they may be extremely vulnerable to even small changes in their consumption (that is, they are likely to be more risk averse than the rich).[61] This is consistent with the large and growing literature which shows that the poor enter into many sophisticated arrangements in order to limit their risk exposure.[62] As noted by Newman (1995), this is not inconsistent with the possibility that the poor are treated better than the rich in the optimal insurance contract.[63] The reason comes from precisely the fact that the poor have more to lose. This means the threat of even a small loss can give good incentives to the poor, making it easier to give them good insurance as well. Newman goes on to show that under reasonable assumptions, this effect can be so strong that the poor would be prepared to take on risky and profitable projects that the rich will avoid. This is, of course, quite different from the predictions of the credit market model—there, it is the rich who do the profitable projects. It is clearly a force towards convergence, and

[60] This phenomenon is well-known in the empirical literature on tenancy and goes under the name of "Tenancy Ladder".

[61] See, for example, Deaton (1989).

[62] See, for example, Udry (1994) and Townsend (1995).

[63] In the sense that the rich would want to have the insurance contract that the poor get in equilibrium.

explains why there is no poverty trap in the model of Banerjee and Newman (1991) on risk bearing. Of course, this kind of result depends crucially on the setup. We would get the opposite result if the insurance market was entirely absent, perhaps because the fixed costs of enforcing such a contract are too high.[64] Then there could easily be a poverty trap—in such a model the poor would underinvest because investment is risky and they are unwilling to bear any risk.[65]

The contrast between the credit market case and the insurance case does not end here. Keeping the insurance contract fixed, increasing risk exposure (by weakening social protection, for example) and increased risk aversion hurt the poor in the insurance context but may actually help them get more credit (since it makes it easier for the lender to "punish" them for defaulting).

The source of this conflict lies in the basic premises of these two narratives: in the insurance market view, the emphasis is on the vulnerability of the poor, i.e., on the fact that they cannot afford any losses; while in the credit market view, the poor are seen as unreliable because they have too little to lose and therefore cannot be punished for defaulting. These views are not necessarily inconsistent: For example, because it may be very costly for the lender to inflict losses on the poor, even though losses hurt the poor a lot.[66] Or the fact that the poor have too little to lose, and therefore unable to invest, might make the sightly less poor extremely averse to the risk of becoming poor and therefore *unwilling to invest*.[67] In other words, it is possible that the credit market and insurance market views reinforce each other. But it remains that the tension between them is real and far from being resolved.

Contracting in product markets is less studied in the context of LDCs. However, as the share of quality-sensitive products in world demand grows (as it has in recent years because of the growth in the new economy), sellers in less developed countries will have to be able to assure buyers that they are getting the desired quality. Since quality is not easy to contract on *ex ante*, this will raise the importance of appropriate contract design in product markets for LDCs. It will also make reputation and brand names much more valuable, making it harder for new entrants. One case where this

[64] See Kanbur (1979) and Kihlstrom and Laffont (1979) for results of this type.

[65] For a simple model of a poverty trap of this type see Banerjee (2000). Morduch (1995) and Walker and Ryan (1990) provide some suggestive evidence for the view that the poor are discouraged by risk from taking up the most profitable opportunities.

[66] See Banerjee (2000) for a more elaborate discussion of these issues.

[67] As in Banerjee (2000).

has already happened is the Indian software industry, where Banerjee and Duflo (2000) show that the more reputed firms both get better contracts and more rewarding projects.

6 The View of Policy

If the evidence and arguments listed above do one thing, it is to challenge the hegemony of the complete markets Arrow-Debreu model as the basis for policy thinking. The usual view in economics seems to be that the complete markets model provides the natural framework, modified perhaps by acknowledging some limited role for transactions costs. My view is that there are many important contexts where these transaction costs are so large and the consequent deviation from the complete markets model so glaring, that it is better to abandon the complete markets model, except in as much as it provides a useful intellectual point of reference. To take an example, in the market studied by Aleem, the transaction cost of about 50 cents on a dollar clearly swamps the interest rate paid to the ultimate lender (the depositor in bank, who gets 10 cents). But this is only the observable part of the transaction cost. Then there is the cost of missed opportunities, because in addition to these high rates the lenders probably impose credit limits and, moreover, some people are completely excluded from the market because no one knows or trusts them.[68] Then there are dynamic costs: The fact that the current borrower underinvests (or earns low net returns on his investment) means that his son/daughter will also be poor and will also not be able to take advantage of his/her opportunities and talents. Finally, there are general equilibrium effects: Inefficiency in investment means low wages and slow capital accumulation today, both of which contribute to poverty and inefficiency tomorrow.

The rejection of the complete markets model should not be seen as a justification for old-fashioned dirigiste polices. There are, of course, good reasons to worry about the deliberate misuse of these policies. But perhaps equally importantly, the recognition that markets often fail does not automatically imply that we should pursue anti-market policies. To take an example, trade protection may be particularly bad if capital markets are imperfect because it reinforces the capital market frictions that slow down the flow of capital toward its best possible uses.

[68] There is, obviously a trade-off here. Some lenders may limit themselves to borrowers they know very well, which brings down the direct transaction cost but increases the costs coming from exclusion and missed opportunities.

What does emerge from the analysis above is the need to build policies that recognize the market failures that are most important in the particular context. Thus, assessing the growth impact of trade policies without taking account of the distributional impact of these policies is self-defeating, since the distributional impact will frame the future pattern of investment.

The imperative of taking specific market failures seriously when making policy is obviously rather vague. To give it some more content, I will now discuss some simple examples of how thinking about the world in this way feeds into specific policy recommendations.

The first example, already alluded to, comes from credit and insurance markets. As we saw, the implications of better social protection tend to be very different in models of credit and insurance. This needs to be taken into account in designing social protection mechanisms.

Second, it is clearly important to try to reduce the cost of credit to the poor. One idea that has received a lot of currency is to make use of peer monitoring and screening by peers through micro-finance institutions.[69] This has two related advantages. First, members of one's peer group may be better at monitoring and screening—this reduces the cost of monitoring. Second, the usual arrangements involve mutual monitoring on a *quid pro quo* basis. Therefore, the interest rate does not have to be raised in order to pay for the cost of monitoring. Lower interest rates, as we already saw, generate better incentives and therefore less monitoring is needed.[70] Other possible interventions include trying to develop systems of credit rating and centralizing credit histories so that the credit markets become less segmented and borrowers have access to the cheapest sources of credit. Developing better systems for recording of property ownership (so that the assets that the poor have can be used as collateral) and a court system that resolves property disputes quickly and effectively (so that lenders believe that they can collect on the collateral), will also help the poor. Helping the poor to develop credit histories and helping them to learn to deal with the financial system more generally (by keeping proper accounts, for example) is yet another potentially fruitful avenue.[71]

Third, giving a central role to issues of credit access gives a new urgency

[69] See Banerjee, Besley, Guinanne (1994) for a model of lending based on peer monitoring and Ghatak (1998) for a model of screening by peers.

[70] This argument relies on the assumption that the time spent on monitoring has no alternative cash-generating use.

[71] This is one way to interpret the contribution of many of the NGOs that work in micro-credit but focus on lending to individuals rather than groups. BRI in Indonesia is a well-known example.

to the old policy prescription of encouraging savings. It is easy to see that one way to get out poverty traps is to raise the savings rate. In Figures 1, 2, 3 and 4, this pushes up the map from current wealth to future wealth and thereby makes poverty traps (both individual and collective) less likely. More generally, as already discussed, the poor may save too little (relative to the social optimum) when capital markets are imperfect, both because the rates paid to depositors tend to be too low (because lending is difficult) and because they do not have the wherewithal to become investors and therefore put their savings to the best possible use. This is compounded by the fact that intermediaries in the credit market may bypass them for the simple reason that given the small volume of their savings, the fixed cost of collecting savings may swamp any potential returns from investing them. Subsidizing access to savings opportunities may therefore be a powerful weapon for helping the poor.

Fourth, thinking of the underlying contract theory clearly gives us a very different perspective on land reform and tenancy reform. In particular, it tells us that it may be possible to achieve many of the desirable productivity effects of such reforms without actually changing any land rights. One example of such an intervention would be a program that improves the outside options of the tenants, such as an employment guarantee scheme.[72] As the tenant's outside option improves, he will be rewarded more and that can lead to an improvement in his incentives.[73]

Finally, if our prediction about the increasing importance of product market contracting in developing countries is borne out, some of these countries will have to make significant policy changes in order to not be left out. In part, this will involve strengthening the court system, but in part also it will require other innovations such as helping domestic companies build a reputation or making it possible for domestic producers to enter into strategic partnerships with reputable MNCs in order to benefit from their reputation.[74]

[72] The strategy of improving outside options is called an empowerment strategy in Banerjee, Gertler and Ghatak (2000), who also provide a formal analysis of how these strategies work.

[73] For a more detailed analysis of land reforms and alternatives to land reform, see Banerjee (1999).

[74] In many cases this may not involve much more than allowing foreign equity participation in domestic companies. However, in cases where countries suffer from a collective reputation problem, there may be a case for a more substantial intervention: penalizing domestic companies that fail to meet some quality standard may be one way of achieving this (see Tirole, 1996, for a model of collective reputation).

7 Conclusion

I have made the choice in this survey of focusing on one rather specific topic in order to explain better the logic of how research in this area of development economics has evolved over the last twenty years, and to pinpoint key gaps in our knowledge in this relatively well-studied area. To conclude, let me now say something about the broader agenda.

The one most important limitation of my survey is its focus on the level, distribution and growth of output as the main outcome of interest. Development economics is much broader: there is a large and important literature which attempts to explain the existence and persistence of institutions such as sharecropping, the village money-lender, ROSCAs, community-based lending networks, cooperatives, etc., based on contract theoretic arguments.[75] There is also a long tradition which argues that institutions act as an independent force in the economy and influence economic outcomes. The Lewis model, which argued that the particular structure of the implicit contract in family farms restrains migration and growth, is perhaps the most well-known example of a model of this type. While there was a period when economists seemed to take the extreme Coasian view that inefficient institutions should not exist, there is now a clearer understanding that there is no good reason why institutions should be set up with an eye towards global optimality. Banerjee and Newman (1998) present an example where a locally efficient institution leads to inefficient global outcomes.[76]

What are still rare are dynamic models where the institutions themselves evolve. Banerjee and Newman (1993) present one such model, aimed at explaining the evolution of the economic institutions of capitalism (large capitalist firms as against self-employment) as a result of change in the wealth distribution.[77] Greenwood and Jovanovic (1990) present a model where the financial sector evolves with growth.

Even rarer are models where institutions not only evolve but actually have feedback effects on the rest of the economy. Some recent examples include Acemoglu and Robinson (1998), which looks at the evolution of the franchise, Acemoglu and Zilibotti (1997), which focuses on the financial sector, and Banerjee and Newman (1998), which looks at the evolution of the

[75] Bardhan (1989) and Greif (1997) are two important sources for this literature. Legros and Newman (1996), Wells (1999), and Prescott and Townsend (2000) are interesting recent examples of this style of work.

[76] Douglas North has been the most influential exponent of this general position in recent years. See Grief (1994) for an interesting and nuanced statement of this view.

[77] See also the related work by Lloyd-Ellis and Bernhardt (2000).

modern sector. But there are not many others, even though it seems clear that this is the process that development economists would like to capture.

A lot has been done in the last twenty years. If that has achieved anything, it is to make us aware of where we would like to be and what needs to be done. Now, I hope, it is only a matter of time.

References

Acemoglu, D. and J. Robinson (1998), "Why Did the West Extend the Franchise? Democracy, Inequality and Growth in Historical Perspective," Centre for Economic Policy Research Discussion Paper No. 1797.

Acemoglu, D. and F. Zilibotti (1997), "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth," *Journal of Political Economy*, Vol. 105 (4), pp. 709-51.

Aghion, P., P. Bacchetta and A. Banerjee (1999), "Capital Markets and the Instability of Open Economies,"unpublished, University College of London and Harvard University.

Aghion, P., A. Banerjee and T. Piketty (1999), "Dualism and Macroeconomic Volatility," *The Quarterly Journal of Economics*, Vol. 114 (4), pp. 1359-97.

Aghion, P. and P. Bolton (1997), "A Theory of Trickle-Down Growth and Development", *The Review of Economic Studies*, Vol. 64 (2), pp. 151-72.

Aleem, I. (1990) "Imperfect Information, Screening and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan", *World Bank Economic Review*, Vol. 3, pp. 329-349.

Angrist, J. and A. Krueger (1999), "Empirical Strategies in Labor Economics", in *Handbook of Labor Economics, Volume 3,* Ashenfelter, O. and D. Card (eds.), Amsterdam; New York: Elsevier Scientific, pp1277-1366.

Banerjee, A. (1992), "A Simple Model of Herd Behavior", *Quarterly Journal of Economics*, Vol. 107 (3), pp. 797-817.

Banerjee, A. (1999), "Prospects and Strategies for Land Reforms," mimeo, MIT.

Banerjee, A. (2000), "The Two Poverties", *Nordic Journal of Political Economy*, Vol. 26 (2), pp. 129-41.

Banerjee, A. (2001), "What Do We Know about Credit Markets", manuscript in preparation, MIT.

Banerjee, A., T. Besley and T. Guinnane (1994), "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test," *The Quarterly Journal of Economics*, Vol. 109 (2), pp. 491-515.

Banerjee, A. and E. Duflo (1999), "Inequality and Growth: What Can the Data Say?," mimeo, MIT.

Banerjee, A. and E. Duflo (2000), "Reputation Effects and the Limits of Contracting," *Quarterly Journal of Economics*, Vol. 115 (3), pp. 989-1017.

Banerjee, A. and E. Duflo (2001), "The Nature of Credit Constraints: Evidence from Indian Bank," mimeo, MIT.

Banerjee, A., P. Gertler and M. Ghatak (2000), "Empowerment and Efficiency: Tenancy Reform in West Bengal," forthcoming, *Journal of Political Economy.*

Banerjee, A. and K. Munshi (2001), "How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur," mimeo, MIT.

Banerjee, A. and A. Newman (1991), "Risk-Bearing and the Theory of Income Distribution," *The Review of Economic Studies*, Vol. 58 (2), pp. 211-35.

Banerjee, A. and A. Newman (1993), "Occupational Choice and the Process of Development," *Journal of Political Economy*, Vol. 101 (2), pp. 274-98.

Banerjee, A. and A. Newman (1994), "The Role of Poverty in Development Theory", *American Economic Review Papers and Proceedings*, May, pp. 211-215.

Banerjee, A. and A. Newman (1998), "Information, the Dual Economy, and Development," *The Review of Economic Studies*, Vol. 65 (4), pp 631-53.

Bardhan, Pranab (1983), "Labor-Tying in a Poor Agrarian Economy: A Theoretical and Empirical Analysis", *Quarterly Journal of Economics,* Vol 98 (3), pp *501-514.*

Bardhan, Pranab, ed. (1989). *The economic theory of agrarian institutions*. Oxford; New York; Toronto and Melbourne: Oxford University Press.

Benabou, R. (1996), "Inequality and Growth," in *NBER Macroeconomics Annual 1996.* Cambridge: MIT Press, pp 11-74.

Besley, T. and A. Case (1994), "Diffusion as a Learning Process: Evidence from HYV Cotton", Princeton, Research Program in Development Studies Discussion Paper #174.

Besley, T., S. Coate and G. Loury (1993), "The Economics of Rotating Savings and Credit Associations," *American Economic Review*, Vol. 83 (4), pp. 792-810.

Bhaduri, A. (1977), "On the Formation of Usurious Interest Rates in Backward Agriculture," *Cambridge Journal of Economics*, Vol. 1 (4), pp. 341-52.

Bhagwati, J. (1982), "Directly Unproductive, Profit-seeking (DUP) Activities," *Journal of Political Economy*, Vol. 90 (5), pp. 988-1002.

Bottomley, A. (1963) "The Cost of Administering Private Loans in Underdeveloped Rural Areas," *Oxford Economic Papers*, Vol. 15 (2), pp. 154-163.

Bowles, Samuel and Herbert Gintis (2001), "The Intergenerational Transmission of Economic Status: Education, Class, and Genetics," in M. Feldman (ed.) *Genetics, Behavior and Society, a volume in Neil Smelser and Paul Baltes, eds., International Encyclopedia of the Social and Behavioral Sciences.* Oxford: Elsevier (forthcoming).

Caballero, R. and M. Hammour (2000), "Creative Destruction in Development: Institutions, Crises, and Restructuring," mimeo, MIT

Card, D. and A. Krueger (1992), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, Vol. 100 (1), pp. 1-40.

Case, Anne and Angus Deaton, "School Inputs and Educational outcomes in South Africa", Vol. 114(3), pp1047-1084.

Casselli, F. and J. Ventura (1996), "A Representative-Agent Theory of Income Distribution", mimeo, MIT.

Cheung, Steven N. S.(1968), "Private Property Rights and Sharecropping," *Journal of Political Economy*, Vol. 76 (6), pp 1107-1122.

Cole, H., G. Mailath and A. Postlewaite (1992), "Social Norms, Savings Behavior, and Growth," *Journal of Political Economy*, Vol. 100 (6), pp. 1092- 1125.

Cunat, Vicente, (2000) "Trade Credit: Suppliers as Debt Collectors and Insurance Providers", LSE-FMG discussion paper.

Das-Gupta, A. *Reports on Informal Credit Markets in India: Summary.* New Delhi: National Institute of Public Finance and Policy, 1989.

Dasgupta, P. (1993), *An Inquiry into Well-Being and Destitution*, Oxford: Oxford University Press.

Dasgupta, P. (1997) "Poverty Traps," in *Advances in Economics and Econometrics: Theory and applications. Seventh World Congress. Volume 2.* Kreps, D. and K. Wallis (eds.), Econometric Society Monographs, No. 27. Cambridge; New York and Melbourne: Cambridge University Press, pp 114-159.

Dasgupta, P. and D. Ray (1986), "Inequality as a Determinant of Malnutrition and Unemployment: Theory", *Economic Journal*, Vol. 96, pp 1011-1034.

Deaton, Angus (1989), "Saving in Developing Countries: Theory and Review", in *Proceedings of the World Bank Annual Conference on Development Economics 1989.* Fischer, S. and D. de Tray (eds.), Washington, D.C.: World Bank, pp 61-96.

Diamond, Douglas (1989) "Reputation Acquisition in Debt Markets" *Journal of Political Economy*, Vol. 97(4), pp. 828-62.

Duflo, Esther (2000), "Grandmothers and Granddaughters: Old Age Pension and Intra-Household Allocation in South Africa," NBER Working Paper #8061.

Duflo, Esther (2001), "Schooling and Labor market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*, 91(4), 795-813.

Durlauf, Steven. (1996), "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, Vol. 1 (1), pp. 75-93.

Evans, D. and B. Jovanovic (1989), "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, Vol. 97 (4), pp 808-27.

Fafchamps, Marcel (2000), "Ethnicity and Credit in African Manufacturing," *Journal of Development Economics*, Vol. 61, pp. 205-235.

Fazzari, S., G. Hubbard and B. Petersen (1988), "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, Vol. 0 (1), pp 141-195.

Fazzari, S. and B. Petersen (1993), "Working Capital and Fixed Investment: New Evidence on Financial Constraints," *Rand Journal of Economics*, Vol. 24 (3), pp 328-342.

Galor, O. and J. Zeira (1993), "Income Distribution and Macroeconomics", *Review of Economic Studies*, Vol. 60 (1), pp 35-52.

Ghatak, Maitreesh (2000), "Screening by the Company You Keep: Joint Liability Lending and the Peer Selection Effect," *Economic Journal*, Vol. 110 (465), pp. 601-31.

Ghatak, Subrata. *Rural Money Markets in India*. Delhi, India: MacMillan Company of India, 1976.

Ghatak, M., M. Morelli and T. Sjostrom (2000), "Dynamic Incentives Occupational Mobility, and the American Dream", mimeo.

Ghate, Prabhu. *Informal Finance: Some Findings from Asia*. Oxford; New York; Toronto and Hong Kong: Oxford University Press for the Asian Development Bank, 1992.

Gill, A. and U. C. Singh (1997), "Financial Sector Reforms, Rate of Interest and the Rural Credit Markets: The Role of Informal Lenders in Punjab," *Indian Journal of Applied Economics*, Vol. 6 (4), pp. 37-65.

Greenwood, J. and B. Jovanovic (1990), "Financial Development, Growth, and the Distribution of Income", *Journal of Political Economy,* Vol. 98 (5), pp 1076-1107.

Greif, A. (1994), "Cultural Beliefs and the Organization of Society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of Political Economy,* Vol. 102(5), pp 912-950.

Greif, A. (1997) "Microtheory and Recent Developments in the Study of Economic Institutions through Economic History," in *Advances in Economics and Econometrics: Theory and applications. Seventh World Congress. Volume 2.* Kreps, D. and K. Wallis (eds.), Econometric Society Monographs, No. 27. Cambridge; New York and Melbourne: Cambridge University Press, pp 79-113.

Hart, O. and J. Moore (1994), "A Theory of Debt Based on the Inalienability of Human Capital," *Quarterly Journal of Economics*, Vol. 109 (4), pp. 841-79.

Holmstrom, B. and J. Tirole (1996), "Modeling Aggregate Liquidity," *American Economic Review*, Vol. 86 (2), pp. 187-91.

Irfan, M., et al. *The structure of informal credit market in Pakistan. Research Report, no. 168.* Islamabad: Pakistan Institute of Development Economics, 1999.

Jaffee, D. & T. Russell (1976), "Imperfect Information, Uncertainty, and Credit Rationing," *Quarterly Journal of Economics*, Vol. 90 (4), pp. 651-66.

Jeong, H. and R. Townsend (2000), "An Evaluation of Models of Growth and Inequality", mimeo, University of Chicago.

Johnson, D. G. (1950), "Resource allocation under share contracts, " *Journal of Political Economy*, Vol. 58 (2), pp 111-123 .

Kanbur, S. (1979), "Of Risk Taking and the Personal Distribution of Income", *Journal of Political Economy*, Vol. 87, pp 769-797.

Kihlstrom, R. and J. Laffont (1979), "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion", *Journal of Political Economy*, Vol. 87, pp 719-48.

Kiyotaki, N. and J. Moore (1997), "Credit Cycles," *Journal of Political Economy*, Vol. 105 (2), pp. 211-248.

Kremer, M. (1993), "The O-Ring Theory of Economic Development", *Quarterly Journal of Economics*, Vol. 108 (3), pp 551-575.

Krueger, A., (1974), "The Politucal Economy of the Rent-seeking Society", *American Economic Review,* Vol. 64 (3), pp. 291-303.

Legros, P. and A. Newman (1996), "Wealth Effects, Distribution, and the Theory of Organization", *Journal of Economic Theory,* Vol. 70 (2), pp 312-41.

Lehnert, A. (1998), "Asset Pooling, Credit Rationing, and Growth," Finance and Economic Discussion Paper 1998-52, Board of Governors of the Federal Reserve System.

Lehnert, A., E. Ligon and R. Townsend (1999), "Liquidity Constraints and Incentive Contracts," Macroeconomic Dynamics, Vol. 3 (1), pp. 1-47.

Lewis, W.A. (1954), "Economic Development with Unlimited Supplies of Labor," *The Manchester School*, Vol. 22, pp. 139-191.

Lloyd-Ellis, H. and D. Bernhardt (2000), "Enterprise, Inequality and Economic Development", *The Review of Economic Studies,* Vol. 67 (1), pp 147-68.

Loury, G. (1981), "Intergenerational Transfers and the Distribution of Earnings," *Econometrica*, Vol. 49 (4), pp. 843-67.

Matsuyama, K. (2000), "Endogenous Inequality," Review of Economic Studies, Vol. 67 (4), pp. 743-59.

McMillan, J. and C. Woodruff (1999), "Interfirm Relationships and Informal Credit in Vietnam," *Quarterly Journal of Economics*, Vol. 114 (4), pp. 1285-1320.

Moav, Omer (1999), "Income Distribution and Macroeconomics: Convex Technology and the Role of Intergenerational Transfers," mimeo, Hebrew University.

Mookherjee, D. and D. Ray (2000), "Persistent Inequality", mimeo, Boston University.

Morduch, J. (1995), "Income Smoothing and Consumption Smoothing" *Journal of Economic Perspectives*, Vol. 9 (3), pp 103-14.

Munshi, Kaivan (2000), "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," mimeo, University of Pennsylvania.

Murphy, K., A. Shleifer and R. Vishny (1989), "Industrialization and the Big Push," *Journal of Political Economy*, Vol. 97 (5), pp 1003-26.

Murshid, K. (1992), "Informal Credit Markets in Bangladesh Agriculture: Bane or Boon?" in *Sustainable Agricultural Development: The role of international cooperation: Proceedings of the 21st International Conference of Agricultural Economists, held at Tokyo, Japan, 22-29 August 1991.* Aldershot, U.K.: Dartmouth; Brookfield, Vt: Ashgate.

Newman, A. (1995), "Risk-Bearing and 'Knightian' Entrepreneurship", mimeo, Columbia University.

Nurske, R. (1953), *Problems of Capital Formation in Underdeveloped Countries.* New York: Oxford University Press.

Paulson, A. and R. Townsend (2000), "Entrepreneurship and Liquidity Constraints in Rural Thailand", mimeo.

Piketty, T. (1997), "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing", *The Review of Economic Studies*, Vol. 64 (2), pp 173-89.

Psacharpoulos, G. (1994), "Returns to Investment in Education: A Global Update", *World Development,* Vol. 22(9), pp1325-1343.

Prescott, E. and R. Townsend (2000), "Inequality, Risk Sharing, and the Boundaries of Collective Organizations," mimeo.

Rosenstein-Rodan, P. (1943), "Problems of Industrialization of Eastern and South-eastern Europe," *Economic Journal*, Vol. 53, pp 202-211.

Rutherford, S. (1999), "The Poor and Their Money: An essay about financial services for poor people", mimeo.

Schultz, T. (1964), *Transforming Traditional Agriculture.* New Haven: Yale University Press.

Srinivasan, T. N. (1994), "Destitution: A Discourse," *Journal of Economic Literature*, Vol. 32 (4), pp. 1842-55.

Stiglitz, J. (1974), "Incentives and Risk Sharing in Sharecropping," *The Review of Economic Studies*, Vol. 41 (2), pp 219-55.

Stiglitz, J. (1990), "Peer Monitoring and Credit Markets," *World Bank Economic Review*, Vol. 4 (3), pp 351-66.

Stiglitz, J. and A. Weiss (1981), "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, Vol. 71 (3), pp 393-410.

Strauss, John (1986), "Does Better Nutrition Raise Farm Productivity", *Journal of Political Economy, Vol. 94(2), pp. 297-320.*

Strauss, John and Duncan Thomas (1993) "Human Resources: Empirical Modeling of Household and Family Decisions", in *Handbook of Development Economics,* (ed) T.N. Srinivasan and Jere Behrman, North-Holland Elsevier.

Swaminathan, Madhura (1991), "Segmentation, Collateral Undervaluation, and the Rate of Interest in Agrarian Credit Markets: Some Evidence from Two Villages in South India," *Cambridge Journal of Economics*, Vol. 15, pp. 161-178.

Timberg, Thomas and C.V. Aiyar (1984), "Informal Credit Markets in India", *Economic Development and Cultural Change*, Vol. 33 (1), pp 43-59.

Tirole, Jean (1996), "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)," *Review of Economic Studies*, Vol. 63 (1), pp. 1-22.

Townsend, R. (1995), "Financial Systems in Northern Thai Villages", *The Quarterly Journal of Economics*, Vol. 110 (4), pp 1011-46.

Udry, Christopher (1994). "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria," *Review of Economic Studies*, Vol. 61(3), pp 495-526.

Walker, T. and J. Ryan (1990), *Village and Household Economies in India's Semi-arid Tropics*, Baltimore and London: Johns Hopkins University Press.

Wells, Robin (1999), "Information, Authority and Internal Governance of the Firm," mimeo, Princeton.

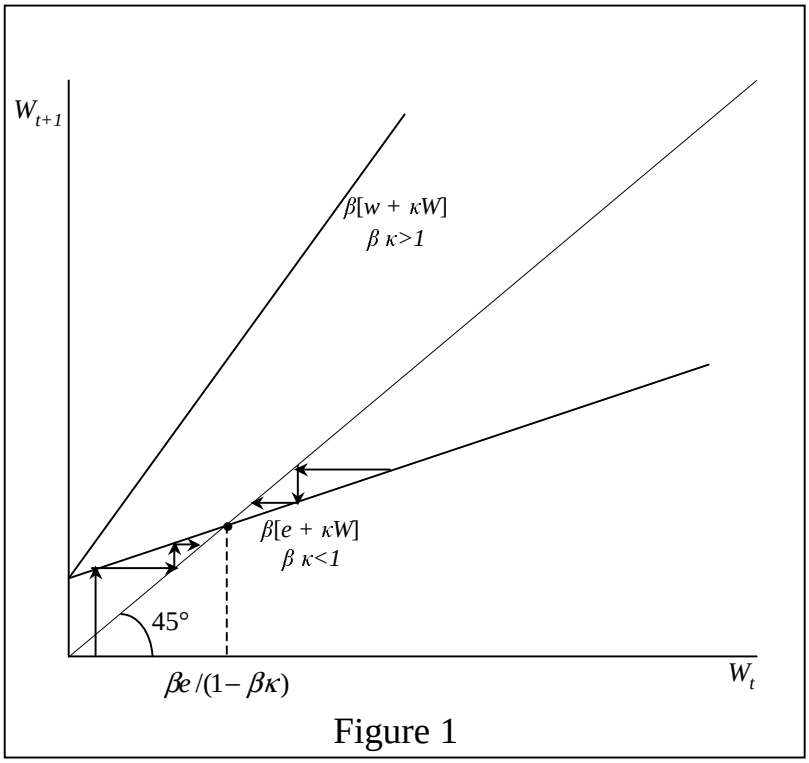

Figure 1

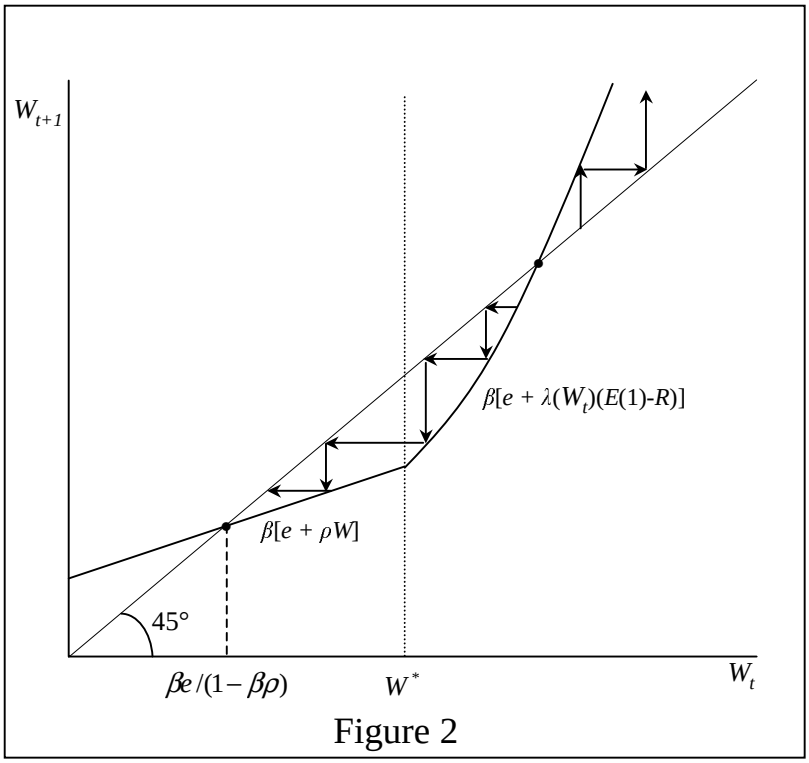

Figure 2

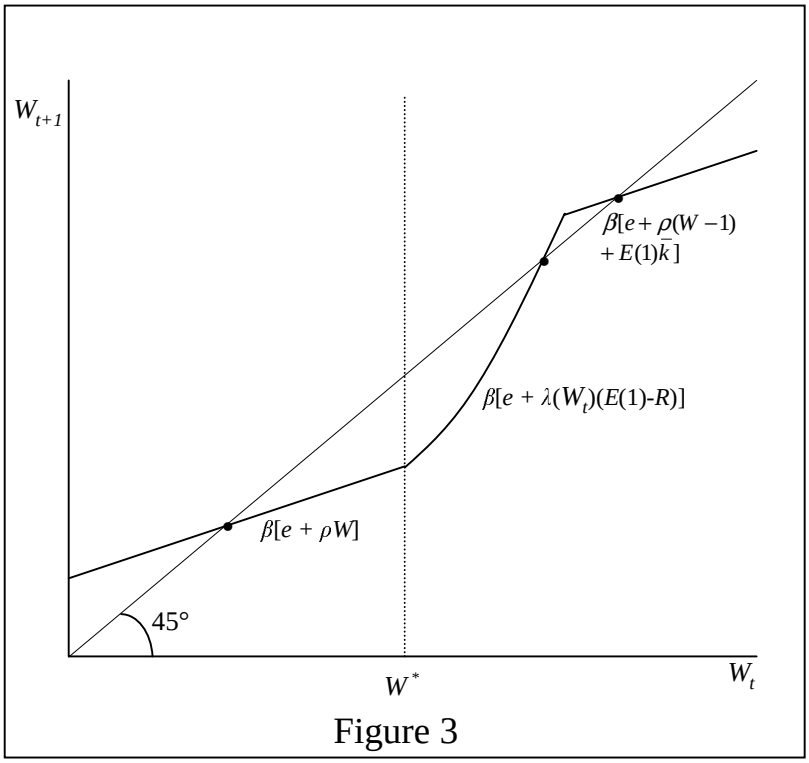

Figure 3

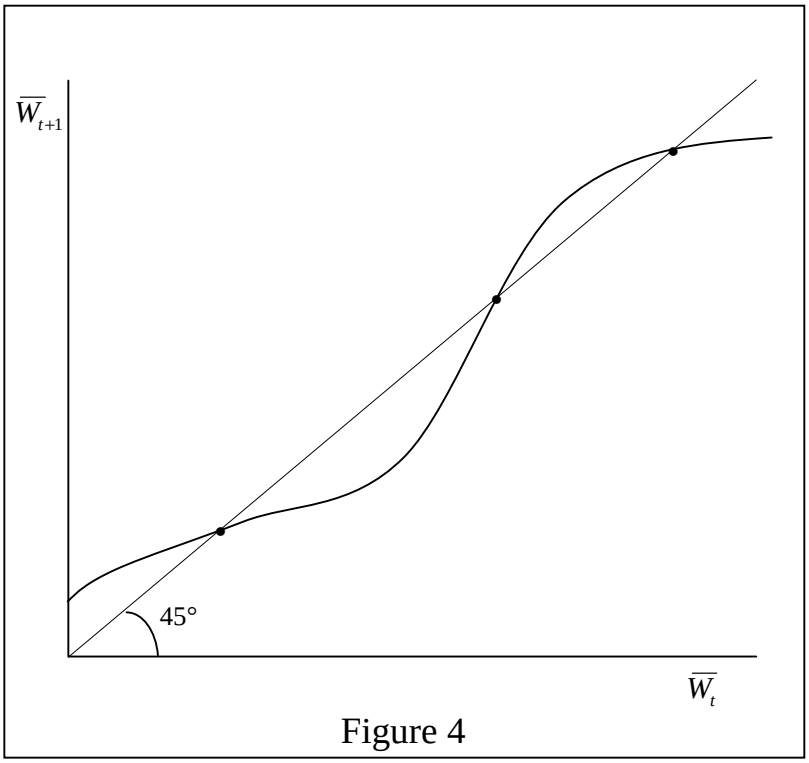

Figure 4