Massachusetts Institute of Technology
Department of Economics
Working Paper Series

GROWTH THEORY THROUGH THE LENS OF DEVELOPMENT ECONOMICS

Abhijit V. Banerjee
Esther Duflo

Working Paper 05-01
December 2004

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://ssrn.com/abstract=651483

Growth Theory through the Lens of Development Economics

Abhijit Banerjee and Esther Duflo
Massachusetts Institute of Technology

Abstract

Growth theory traditionally assumed the existence of an aggregate production function, whose existence and properties are closely tied to the assumption of optimal resource allocation within each economy. We show extensive evidence, culled from the micro-development literature, demonstrating that the assumption of optimal resource allocation fails radically. The key fact is the enormous heterogeneity of rates of return to the same factor within a single economy, a heterogeneity that dwarfs the cross-country heterogeneity in the economy-wide average return. *Prima facie*, we argue, this evidence poses problems for old and new growth theories alike. We then review the literature on various causes of this misallocation. We go on to calibrate a simple model which explicitly introduces the possibility of misallocation into an otherwise standard growth model . We show that, in order to match the data, it is not enough to have misallocated factors: there also needs to be important fixed costs in production. We conclude by outlining the contour of a possible non-aggregate growth theory, and review the existing attempts to take such a model to the data.

JEL numbers O0, O10, O11, O12, O14, O15, O16, O40

Keywords: Non-aggregative growth theory; aggregate production function; factor allocation; non-convexities.

Growth Theory through the Lens of Development Economics

Abhijit V. Banerjee and Esther Duflo*

December 2004

1 Introduction: Neo-classical Growth Theory

The premise of neo-classical growth theory is that it is possible to do a reasonable job of explaining the broad patterns of economic change across countries, by looking at it through the lens of an aggregate production function. The aggregate production function relates the total output of an economy (a country, for example) to the aggregate amounts of labor, human capital and physical capital in the economy, and some simple measure of the level of technology in the economy as a whole. It is formally represented as $\mathcal{F}(A, \overline{K}_P \overline{K}_H, \overline{L})$ where $\overline{K}_P$ and $\overline{K}_H$ are the total amounts of physical and human capital invested, $\overline{L}$ is the total labor endowment of the economy and A is a technology parameter.

The aggregate production function is not meant to be something that physically exists. Rather, it is a convenient construct. Growth theorists, like everyone else, have in mind a world where production functions are associated with people. To see how they proceed, let us start with a model where everyone has the option of starting a firm, and when they do, they have access to an individual production function

$$Y = F(K_P, K_H, L, \theta), \tag{1}$$

where K_P and K_H are the amounts of physical and human capital invested in the firm and L is the amount of labor. θ is a productivity parameter which may vary over time, but at any point of time is a characteristic of the firm's owner. Assume that F is increasing in all its inputs. To make life simpler, assume that there is only one final good in this economy and physical capital is made from it. Also assume

*MIT, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. banerjee@mit.edu, eduflo@mit.edu. For financial support, the authors are grateful to the National Science Foundation under the grant SES- 0137015 (Banerjee), the Alfred P. Sloan Foundation (Duflo) and the John D. and Catherine MacArthur Foundation. We are also grateful to Pranab Bardhan, Michael Kremer, Rohini Pande, Chris Udry and Ivan Werning for helpful conversations, to Philippe Aghion and Seema Jayachandran for detailed comments, and to Charles Cohen and Thomas Wang for excellent research assistance. A part of this material was presented as the Kuznets Memorial Lecture, 2004, at Yale University. We are grateful for the many comments that we received from the audience.

that the population of the economy is described by a distribution function $G_t(W, \theta)$, the joint distribution of W and θ, where W is the wealth of a particular individual and θ is his productivity parameter. Let $\widetilde{G}(\theta)$ be the corresponding partial distribution on θ.

The lives of people, as often is the case in economic models, is rather dreary: In each period, each person, given his wealth, his θ and the prices of the inputs, decides whether to set up a firm, and if so how to invest in physical and human capital. At the end of the period, once he gets returns from the investment and possibly other incomes, he consumes and the period ends. The consumption decision is based on maximizing the following utility function:

$$\sum_{t=0}^{\infty} \delta^t U(C_t, \theta), 0 < \delta < 1. \tag{2}$$

1.1 The Aggregate Production Function

The key assumption behind the construction of the aggregate production function is that all factor markets are perfect, in the sense that individuals can buy or sell as much as they want at a given price. With perfect factor markets (and no risk) the market must allocate the available supply of inputs to maximize total output. Assuming that the distribution of productivities does not vary across countries, we can therefore define $\mathcal{F}(\overline{K}_P, \overline{K}_H, \overline{L})$ to be:

$$\begin{aligned} & \max_{\{K_P(\theta), K_H(\theta), L(\theta)\}} \{\int_\theta F(K_P(\theta), K_H(\theta), L(\theta), \theta) d\widetilde{G}(\theta)\} \\ \text{subject to} \int_\theta K_P(\theta) d\theta \quad = \quad & \overline{K}_P, \int_\theta K_H(\theta) d\theta = \overline{K}_H, and \int_\theta L(\theta) d\theta = \overline{L}. \end{aligned}$$

This is the aggregate production function. It is notable that the distribution of wealth does not enter anywhere in this calculation. This reflects the fact that with perfect factor markets, there is no necessary link between what someone owns and what gets used in the firm that he owns. The fact that $\widetilde{G}(\theta)$ does not enter as an argument of $\mathcal{F}(\overline{K}_P, \overline{K}_H, \overline{L})$ reflects our assumption that the distribution of productivities does not vary across countries.

It should be clear from the construction that there is no reason to expect a close relation between the "shape" of the individual production function and the shape of the aggregate function. Indeed it is well known that aggregation tends to convexify the production set: In other words, the aggregate production function may be concave even if the individual production functions are not. In this environment where there are a continuum of firms, the (weak) concavity of the aggregate production function is guaranteed as long as the average product of the inputs in the individual production functions is bounded in the sense that there is a λ such that $F(\lambda K_P, \lambda K_H, \lambda L, \theta) \leq \lambda \|(K_P, K_H, L, \theta)\|$ for all K_P, K_H, L and θ. It follows that the concavity of the individual functions is sufficient for the concavity of the aggregate but by no

means necessary: The aggregate production would also be concave if the individual production functions were S-shaped (convex to start out and then becoming concave). Alternately, the individual production function being bounded is enough to guarantee concavity of the aggregate production function. Moreover, the aggregate production function will typically be differentiable almost everywhere.

It is a corollary of this result that the easiest way to generate an aggregate production function with increasing returns is to base the increasing returns not on the shape of the individual production function, but rather on the possibility of externalities across firms. If there are sufficiently strong positive externalities between investment in one firm and investment in another, increasing the total capital stock in all of them together will increase aggregate output by more (in proportional terms) than the same increase in a single firm would raise the firm's output, which could easily make the aggregate production function convex. This is the reason why externalities have been intimately connected, in the growth literature, with the possibility of increasing returns.

The assumption of perfect factor markets is therefore at the heart of neo-classical growth theory. It buys us two key properties: The fact that the ownership of factors does not matter, i.e., that an aggregate production function exists; and that it is concave. The next sub-section shows how powerful these two assumptions can be.

1.2 The Logic of Convergence

Assume for simplicity that production only requires physical capital and labor and that the aggregate production function, $\mathcal{F}(\overline{K_p}, \overline{L})$ defined as above, exhibits constant returns and is concave, increasing, almost everywhere differentiable and eventually strictly concave, in the sense that $\mathcal{F}'' < \varepsilon < 0$, for any $\overline{K}_p > \widetilde{\overline{K_p}}$. As noted above, this does not require the individual production functions to have this shape, though it does impose some constraints on what the individual functions can be like. It does however require that the distribution of firm-level productivities is the same everywhere.

Under our assumption that capital markets are perfect, in the sense that people can borrow and lend as much as they want at the common going rate, r_t, the marginal returns to capital must be the same for everybody in the economy. This, combined with the preferences as represented by (2), has the immediate consequence that for everybody in the economy:

$$U'(C_t, \theta) = \delta r_t U''(C_{t+1}, \theta).$$

It follows that everybody's consumption in the economy must grow as long as $\delta r_t > 1$ and shrink if $\delta r_t < 1$. And since consumption must increase with wealth, it follows that everyone must be getting richer if and only if $\delta r_t > 1$, and consequently the aggregate wealth of the economy must be growing as long as $\delta r_t > 1$. In a closed economy, the total wealth must be equal to the total capital stock, and

therefore the capital stock must also be increasing under the same conditions.

Credit market equilibrium, under perfect capital markets, implies that $\mathcal{F}'(\overline{K}_{Pt}, L) = r_t$. The fact that F is eventually strictly concave implies that as the aggregate capital stock grows, its marginal product must eventually start falling, at a rate bounded away from 0. This process can only stop when $\delta\mathcal{F}'(\overline{K}_{Pt}, L) = 1$. As long as the production function is the same everywhere, all countries must end up equally wealthy

The logic of convergence starts with the fact that in poor countries, capital is scarce, which combined with the concavity of the aggregate production function implies that the return on the capital stock should be high. Even with the same fraction of these higher returns being reinvested, the growth rate in the poorer countries would be higher. Moreover, the high returns should encourage a higher reinvestment rate, unless the income effect on consumption is strong enough to dominate. Together, they should make the poorer countries grow faster and catch up with the rich ones.

Yet poorer countries do not grow faster. According to Mankiw, Romer and Weil (1992), the correlation between the growth rate and the initial level of Gross Domestic Product is small, and if anything, positive (the coefficient of the log of the GDP in 1960 on growth rate between 1960 and 1992 is 0.0943). Somewhere along the way, the logic seems to have broken down.

Understanding the failure of convergence has been one of the key endeavors of the economics of growth. What we try to do in this chapter is to argue that the failure of this approach is intimately tied to the failure of the assumptions that underlie the construction of the aggregate production function and to suggest an alternative approach to growth theory that abandons the aggregate production.

We start by discussing, in section 2, the two implications of the neo-classical model that are at the root of the convergence result: Both rates of returns and investment rates should be higher in poor countries. We show that, in fact, neither rates of returns nor investment are, on average, much higher in poor countries. Moreover, contrary to what the aggregate production approach implies, there are large variations in rate of returns within countries, and large variation in the extent to which profitable investment opportunities are taken advantage of.

In section 3, we ask whether the puzzle (of no convergence) can be solved, while maintaining the aggregate production function, by theories that focus on reasons for technological backwardness in poor countries. We argue that this class of explanations is not consistent with the empirical evidence which suggests that many firms in poor countries do use the latest technologies, while others in the same country use obsolete modes of production. In other words, what we need to explain is less the overall technological backwardness and more why some firms do not adopt profitable technologies that are available to them (though perhaps not affordable).

In section 4, we attempt to suggest some answers to the question of why firms and people in devel-

oping countries do not always avail themselves of the best opportunities afforded to them. We review various possible sources of the inefficient use of resources: government failures, credit constraints, insurance failure, externalities, family dynamics, and behavioral issues. We argue that each of these market imperfections can explain why investment may not always take place where the rates of returns are the highest, and therefore why resources may be misallocated within countries. This misallocation, in turn, drives down returns and this may lower the overall investment rate. In section 5, we calibrate plausible magnitudes for the aggregate static impact of misallocation of capital within countries We show that, combined with individual production functions characterized by fixed costs, the misallocation of capital implied by the variation of the returns to capital observed within countries can explain the main aggregate puzzles: the low aggregate productivity of capital, and the low Total Factor Productivity in developing countries, relative to rich countries. Non-aggregative growth models thus seem to have the potential to explain why poor countries remain poor.

The last section provides an introduction to an alternative growth theory that does not require the existence of an aggregate production function, and therefore can accommodate the misallocation of resources. We then review the attempts to empirically test these models. We argue that the failure to take seriously the implications of non-aggregative models have led to results that are very hard to interpret. To end, we discuss an alternative empirical approach illustrated by some recent calibration exercises based on growth models that take the misallocation of resources seriously.

2 Rates of Return and Investment Rates in Poor Countries

In this section, we examine whether the two main implications of the neo-classical model are verified in the data: Are returns and investment rates higher in poor countries?

2.1 Are returns higher in poor countries?

2.1.1 Physical Capital

• Indirect Estimates

One way to look at this question is to look at the interest rates people are willing to pay. Unless people have absolutely no assets that they can currently sell, the marginal product of whatever they are doing with the marginal unit of capital should be no less than the interest rate: If this were not true, they could simply divert the last unit of capital toward whatever they are borrowing the money for and be better off.

There is a long line of papers that describe the workings of credit markets in poor countries (Banerjee

(2003) summarizes this evidence). The evidence suggests that a substantial fraction of borrowing takes place at very high interest rates.

A first source of evidence is the "Summary Report on Informal Credit Markets in India" (Dasgupta (1989)), which 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 six villages in Kerala and Tamil Nadu, carried out by the Centre for Development Studies. The average annual interest rate charged by professional moneylenders (who provide 45.6% of the credit) in these surveys is about 52%. For the urban sector, the data is based on various case surveys of specific classes of informal lenders, many of whom lend mostly to trade or industry. For finance corporations, they report that the minimum lending rate on loans of less than one year is 48%. For hire-purchase companies in Delhi, the lending rate was between 28% and 41%. For auto financiers in Namakkal, the lending rate was 40%. For handloom financiers in Bangalore and Karur, the lending rate varied between 44% and 68%.

Several other studies reach similar conclusions. A study by Timberg and Aiyar (1984) reports data on indigenous-style bankers in India, based on surveys they carried out: 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). 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%. Ghate (1992) reports on a number of case studies from all over Asia: The case study from Thailand found that interest rates were 5-7% per month in the north and northeast (5% per month is 80% per year and 7% per month is 125%). Murshid (1992) studies Dhaner Upore (cash for kind) loans in Bangladesh (you get some amount in rice now and repay some amount in rice later) and reports that the interest rate is 40% for a 3-5 month loan period. The Fafchamps (2000) study of informal trade credit in Kenya and Zimbabwe reports an average monthly interest rate of 2.5% (corresponding to an annualized rate of 34%) but also notes that this is the rate for the dominant trading group (Indians in Kenya, whites in Zimbabwe), while the blacks pay 5% per month in both places.

The fact that interest rates are so high could reflect the high risk of default. However, this does not appear to be the case, since several of studies mentioned above give the default rates that go with these high interest rates. The study by Dasgupta (1989) attempts to decompose the observed interest rates into their various components,[1] and finds that the default costs explain 7 per cent (not 7 percentage points!) of the total interest costs for auto financiers in Namakkal and handloom financiers in Bangalore and Karur, 4% for finance companies and 3% for hire-purchase companies. The same study reports that

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

in four case studies of moneylenders in rural India they found default rates explained about 23% of the observed interest rate. Timberg and Aiyar (1984), whose study is also mentioned above, report that average default losses for the informal lenders they studied ranges between 0.5% and 1.5% of working funds. The study by Aleem Aleem (1990) gives default rates for each individual lender. The median default rate is between 1.5 and 2%, and the maximum is 10%.[2]

Finally, it does not seem to be the case that these high rates are only paid by those who have absolutely no assets left. The "Summary Report on Informal Credit Markets in India" (Dasgupta (1989)) reports that several of the categories of lenders that have already been mentioned, such as handloom financiers and finance corporations, focus almost exclusively on financing trade and industry while Timberg and Aiyar (1984) report that for Shikarpuri bankers at least 75% of the money goes to finance trade and, to lesser extent, industry. In other words, they only lend to established firms. It is hard to imagine, though not impossible, that all the firms have literally no assets that they can sell. Ghate (1992) also concludes that the bulk of informal credit goes to finance trade and production, and Murshid (1992), also mentioned above, 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.

Udry (2003) obtains similar indirect estimates by restricting himself to a sector where loans are used for productive purpose, the market for spare taxi parts in Accra, Ghana. He collected 40 pairs of observations on price and expected life for a particular used car part sold by a particular dealer (e.g., alternator, steering rack, drive shaft). Solving for the discount rate which makes the expected discounted cost of two similar parts equal gives a lower bound to the returns to capital. He obtains an estimate of 77% for the median discount rate.

Together, these studies thus suggest that people are willing to pay high interest rates for loans used for productive purpose, which suggests that the rates of return to capital are indeed high in developing countries, at least for some people.

- Direct Estimates

Some studies have tried to come up with more direct estimates of the rates of returns to capital. The "standard" way to estimate returns to capital is to posit a production function (translog and Cobb-Douglas, generally) and to estimate its parameters using OLS regression, or instrumenting capital with

[2]Here we make no attempt to answer the question of why the interest rates are so high. Banerjee (2003) argues that it is not implausible that the enormous gap between borrowing and lending rates implied by these numbers, simply reflects the cost of lending (monitoring and contracting costs of various kinds). Hoff and Stiglitz (1998) suggest an important role for monopolistic competition, in the presence of a fixed cost of lending. There is also a view that the market for credit is monoploized by a small number of lenders who earn excess profits, but Aleem (1990) finds no evidence of excess profits.

its price. Using this methodology, Bigsten, Isaksson, Soderbom and Al (2000) estimate returns to physical and human capital in five African countries. They estimate rates of returns ranging from 10% to 32%. McKenzie and Woodruff (2003) estimate parametric and non-parametric relationships between firm earnings and firm capital. Their estimates suggest huge returns to capital for these small firms: For firms with less than $200 invested, the rate of returns reaches 15% per *month*, well above the informal interest rates available in pawn shops or through micro-credit programs (on the order of 3% per month). Estimated rates of return decline with investment, but remain high (7% to 10% for firms with investment between $200 and $500, 5% for firms with investment between $500 and $1,000).

Such studies present serious methodological issues, however. First, the investment levels are likely to be correlated with omitted variables. For example, in a world without credit constraints, investment will be positively correlated with the expected returns to investment, generating a positive "ability bias" (Olley and Pakes (1996)). McKenzie and Woodruff attempt to control for managerial ability by including the firm owner's wage in previous employment, but this may go only part of the way if individuals choose to enter self-employment precisely because their expected productivity in self-employment is much larger than their productivity in an employed job. Conversely, there could be a negative ability bias, if capital is allocated to firms in order to avoid their failure.

Banerjee and Duflo (2004) take advantage of a change in the definition of the so-called "priority sector" in India to circumvent these difficulties. All banks in India are required to lend at least 40% of their net credit to the "priority sector", which includes small-scale industry, at an interest rate that is required to be no more than 4% above their prime lending rate. In January, 1998, the limit on total investment in plants and machinery for a firm to be eligible for inclusion in the small-scale industry category was raised from Rs. 6.5 million to Rs. 30 million. In 2000, the limit was lowered back to Rs 10 million Banerjee and Duflo (2004) first show that, after the reforms, newly eligible firms (those with investment between 6.5 million and 30 million) received on average larger increments in their working capital limit than smaller firms. They then show that the sales and profits increased faster for these firms during the same period. The opposite happened when the priority sector was contracted again. Putting these two facts together, they use the variation in the eligibility rule over time to construct instrumental variable estimates of the impact of working capital on sales and profits. After computing a non-subsidized cost of capital, they estimate that the returns to capital in these firms must be at least 74%.

There is also direct evidence of very high rates of returns on productive investment in agriculture. Goldstein and Udry (1999) estimate the rates of returns to the production of pineapple in Ghana. The rate of returns associated with switching from the traditional maize and Cassava intercrops to pineapple is estimated to be in excess of 1,200%! Few people grow pineapple, however, and this figure may hide some heterogeneity between those who have switched to pineapple and those who have not.

Evidence from experimental farms also suggests that, in Africa, the rate of returns to using chemical fertilizer (for maize) would also be high. However, this evidence may not be realistic, if the ideal conditions of an experimental farm cannot be reproduced on actual farms. Foster and Rosenzweig (1995) show, for example, that the returns to switching to high yielding varieties were actually low in the early years of the green revolution in India, and even negative for farmers without an education. This is despite the fact that these varieties had precisely been selected for having high yields, in proper conditions. But they required complementary inputs in the correct quantities and timing. If farmers were not able or did not know how to supply those, the rates of returns were actually low.

To estimate the rates of returns to using fertilizer in actual farms in Kenya, Duflo, Kremer and Robinson (2003), in collaboration with a small NGO, set up small scale randomized trials on people's farms: Each farmer in the trials delimited two small plots. On one randomly selected plot, a field officer from the NGO helped the farmer apply fertilizer. Other than that, the farmers continued to farm as usual. They find that the rates of returns from using a small amount of fertilizer varied from 169% to 500% depending on the year, although of returns decline fast with the quantity used on a plot of a given size. This is not inconsistent with the results in Foster and Rosenzweig (1995), since by the time this study was conducted in Kenya, chemical fertilizer was a well established and well understood technology, which did not need many complementary inputs.

The direct estimates thus tend to confirm the indirect estimates: While there are some settings where investment is not productive, there seems to be investment opportunities which yield substantial rates of returns.

- How high is the marginal product on average?

The fact that the marginal product in some firms is 50% or 100% or even more does not imply that the average of the marginal products across all firms is nearly as high. Of course, if capital always went to its best use, the notion of the average of the marginal products does not make sense. The presumption here is that there may be an equilibrium where the marginal products are not equalized across firms.

One way to get at the average of the marginal products is to look at the Incremental Capital Output Ratio (ICOR) for the country as a whole. The ICOR measures the increase in output predicted by a one unit increase in capital stock. It is calculated by extrapolating from the past experience of the country and assumes that the next unit of capital will be used exactly as efficiently (or inefficiently) as the last one. The inverse of the ICOR therefore gives an upper bound for the average marginal product for the economy—it is an upper bound because the calculation of the ICOR does not control for the effect of the increases in the other factors of production which also contributes to the increase in output.[3] For the

[3]The implicit assumption that the other factors of production are growing is probably reasonable for most developing

late 1990s, the IMF estimates that the ICOR is over 4.5 for India and 3.7 for Uganda. The implied upper bound on the average marginal product is 22% for India and 27% in Uganda. This is also consistent with the work of Pessoa, Cavalcanti-Ferreira and Velloso (2004) who estimate a production function using cross-country data and calculate marginal products for developing countries which are in the 10-20% range. It seems that the average returns are actually not much higher than the 9% or so, which is the usual estimate for the average stock market return in the US.

- Variations in the marginal products across firms.

To reconcile the high direct and indirect estimates of the marginal returns we just discussed and an average marginal product of 22% in India, it would have to be that there is substantial variation in the marginal product of capital within the country. Given that the inefficiency of the Indian public sector is legendary, this may just be explained by the investment in the public sector. However, since the ICOR is from the late 1990s, when there was little new investment (or even disinvestment) in the public sector, there must also be many firms in the private sector with marginal returns substantially below 22%. The micro evidence reported in Banerjee (2004), which shows that there is very substantial variation in the interest rate within the same sub-economy, certainly goes in this direction. The Timberg and Aiyar (1984) study mentioned above, is one source of this evidence: It reports that the Shikarpuri lenders charged rates that were as low as 21% and as high as 120%, and some established traders on the Calcutta and Bombay commodity markets could raise funds for as little as 9%. The study by Aleem (1990), also mentioned above, reports that the standard deviation of the interest rate was 38.14%. Given that the average lending rate was 78.5%, this tells us that an interest rate of 2% and an interest rate of 150% were both within two standard deviations of the mean. Unfortunately, we cannot quite assume from this that there are some borrowers whose marginal product is 9% or less: The interest rate may not be the marginal product if the borrowers who have access to these rates are credit constrained. Nevertheless, given that these are typically very established traders, this is less likely than it would be otherwise.

Ideally we would settle this issue on the basis of direct evidence on the misallocation of capital, by providing direct evidence on variations in rates of return across groups of firms. Unfortunately such evidence is not easy to come by, since it is difficult to consistently measure the marginal product of capital. However, there is some rather suggestive evidence from the knitted garment industry in the Southern Indian town of Tirupur (Banerjee and Munshi (2004); Banerjee, Duflo and Munshi (2003)). Two groups of people operate in Tirupur: the Gounders, who issue from a small, wealthy, agricultural community from the area around Tirupur, who have moved into the ready-made garment industry because there was not much investment opportunity in agriculture. Outsiders from various regions and communities

countries, except perhaps in Africa.

started joining the city in the 1990s. The Gounders have, unsurprisingly, much stronger ties in the local community, and thus better access to local finance, but may be expected to have less natural abilities for garment manufacturing than the outsiders, who came to Tirupur precisely because of its reputation as a center for garment export. The Gounders own about twice as much capital as the outsiders on average. They maintain a higher capital-output ratio than the outsiders at all levels of experience, though the gap narrows over time. The data also suggest that they make less good use of their capital than the outsiders: While the outsiders start with lower production and exports than the Gounders, their experience profile is much steeper, and they eventually overtake the Gounders at high levels of experience, even though they have lower capital stock throughout. This data therefore suggests that capital does not flow where the rates of return are highest: The outsiders are clearly more able than the Gounders, but they nevertheless invest less.[4]

To summarize, the evidence on returns to physical capital in developing countries suggests that there are instances with high rates of return, while the average of the marginal rates of return across firms does not appear to be that high. This suggests a coexistence of very high and very low rates of return in the same economy.

2.1.2 Human Capital

• Education

The standard source of data on the rate of return to education is Psacharopoulos (1973; 1985; 1994; 2002)) who compiles average Mincerian returns to education (the coefficient of years of schooling in a regression of log(wages) on years of schooling) as well as what he call "full returns" to education by level of schooling. Compared to Mincerian returns, full returns take into account the variation in the cost of schooling according to year of schooling: The opportunity cost of attending primary school is low, because 6 to 12 year old children do not earn the same wage as adults; and the direct costs of education increase with the level of schooling.

On the basis of this data, Psacharopoulos argues that returns to education are substantial, and that they are larger in poor countries than in rich countries. We re-examine the claim that returns to education are larger in poor countries, using data on traditional Mincerian returns, which have the advantage of being directly comparable. We start with the latest compilation of rates of returns, available in Psacharopoulos (2002) and on the World Bank web site. We update it as much as possible, using studies that seem to have been overlooked by Psacharopoulos, or that have appeared since then (the updated data set and the references are presented in the appendix).[5] We flag the observations that Bennell (1996)

[4]This is not because capital and talent happen to be substitutes. In this data, as it is generally assumed, capital and ability appear to be complements.

[5]The bulk of the update is for African countries, where Bennell (1996) had systematically investigated the Psacharoupou-

rated as being of "poor" or "very poor" quality. We complete this updated database by adding data on years of schooling for the year of the study when it was not reported by Psacharopoulos.

Using the preferred data, the Mincerian rates of returns seem to vary little across countries: The mean rate of returns is 8.96, with a standard deviation of 2.2. The maximum rate of returns to education (Pakistan) is 15.4%, and the minimum is 2.7% (Italy). Averaging within continents, the average returns are highest in Latin America (11%) and lowest in the Europe and the U.S. (7%), with Africa and Asia in the middle.

If we run an OLS regression of the rates of returns to education on the average educational attainment (number of years of education), using the preferred data (updated database without the low quality data), the coefficient is -0.26, and is significant at 10% level (table 1, column 3). The returns to education predicted from this regression range from 6.9% for the country with the lowest education level to 10.1% for the country with the highest education level. This is a small range (smaller than the variation in the estimates of the returns to education of a single country, or even in different specifications in a single paper!): There is therefore no *prima facie* evidence that returns to education are much higher when education is lower, although the relationship is indeed negative. Columns (1) and (2) in the same table show that the data construction matters: When the countries with "poor" quality are included, the coefficient of years of education increases to -0.45. When only the 38 countries in the latest Psacharopoulos update are included (most countries are dropped because the database does not report years of education, even for countries where it is clearly available–Austria for example), the coefficient more than doubles, to -0.71. On the whole, this strong negative number does appear to be an artifact of data quality.

In column (4), we directly regress the Mincerian returns to education on GDP, and we find a small and significant negative relationship. However, this is counteracted by the fact that teacher salary grows less fast than GDP, and the cost of education is thus not proportional to GDP: In column (5) we regress the log of the teacher salary on the log of GDP per capita.[6] The coefficient is significantly less than one, suggesting that teachers are relatively more expensive in poor countries. This is to some extent attenuated by the fact that class sizes are larger in poor countries (which tends to make education cheaper). We then compute the returns to educating a child for one year as the ratio of the lifetime benefit of one year of education (assuming a life span of 30 years, a discount rate of 5%, a share of wage in GDP of 60%, and no growth), to the direct cost of education (assuming that teacher salary is 85% of the cost of education). In column (6), we regress this ratio on GDP: There is no relationship between this measure of returns and GDP.[7] If we factor in indirect costs (as a fraction of GDP) (in column (7)), the relationship becomes

los data, and found that many of the underlying studies were unreliable.

[6]The teacher salary data is obtained from the "Occupational Wages Around the World" database (Freeman and Oostendorp (2001)).

[7]Note that by assuming that the lifespan is the same in poor and rich countries, we are biasing upwards the returns in

slightly more negative, but still insignificant. On balance, the returns to one more year of education are therefore no higher in poor countries.

• Health

Education is not the only dimension of human capital. In developing countries, investment in nutrition and health has been hypothesized to have potentially high returns at moderate levels of investment. The report of the Commission for Macroeconomics and Health (Commission on Macroeconomics and Health (2001)), for example, estimated returns to investing in health to be on the order of 500%, mostly on the basis of cross-country growth regressions. Several excellent recent surveys (Strauss and Thomas (1995); Strauss and Thomas (1998), Thomas (2001) and Thomas and Frankenberg (2002)) summarize the existing literature on the impact of different measures of health on fitness and productivity, and lead to a much more nuanced conclusion.

There is substantial experimental evidence that supplementation in iron and vitamin A increases productivity at relatively low cost. Unfortunately, not all studies report explicit rates of returns calculations. The few numbers that are available suggest that some basic health intervention can have high of returns: Basta, Soekirman, Karyadi and Scrimshaw (1979) studies an iron supplementation experiment conducted among rubber tree tappers in Indonesia. Baseline health measures indicated that 45% of the study population was anemic. The intervention combined an iron supplement and an incentive (given to both treatment and control groups) to take the pill on time. Work productivity in the treatment group increased by 20% (or $132 per year), at a cost per worker-year of $0.50. Even taking into account the cost of the incentive ($11 per year), the intervention suggests extremely high rates of returns. Thomas, Frankenberg, Friedman, Habicht and Al (2003) obtain lower, but still high, estimates in a larger experiment, also conducted in Indonesia: They found that iron supplementation experiments in Indonesia reduced anemia, increased the probably of participating in the labor market, and increased earnings of self-employed workers. They estimate that, for self-employed males, the benefits of iron supplementation amount to $40 per year, at a cost of $6 per year.[8] The cost benefit analysis of a de-worming program (Miguel and Kremer (2004)) in Kenya reports estimates of a similar order of magnitude: Taking into account externalities (due to the contagious nature of worms), the program led to an average increase in school participation of 0.14 years. Using a reasonable figure for the returns to a year of education, this additional schooling will lead to a benefit of $30 over the life of the child, at a cost of $0.49 per child per year. Not all interventions have the same rates of return however: A study of Chinese cotton mill workers (Li, Chen, Yan, Deurenberg, Garby and Hautvast (1994)) led to a significant increase in fitness, but no

poor countries.

[8]This number takes into account the fact that only 20% of the Indonesian population is iron deficient: The private returns of iron supplementation for someone who knew they were iron deficient–which they can find out using a simple finger prick–would be $200).

corresponding increase in productivity. Likewise, the intervention analyzed by Thomas et al. (2003) had no effect on earnings or labor force participation of women.

In summary, while there is not much debate on the impact of fighting anemia (through iron supplementation or de-worming) on work capacity, there is more heterogeneity amongst estimates of economic rates of return of these interventions. The heterogeneity is even larger when we consider other forms of health interventions, reviewed, for example, in Strauss and Thomas (1995), or when one compares various human capital interventions. As in the case of physical capital, there are instances of high returns, and substantial heterogeneity in returns.

2.1.3 Taking Stock: Returns on Capital

The marginal product of physical and human capital in developing countries seems very high in some instances, but not necessarily uniformly. The average of the marginal products of physical capital in India may be as low as 22%, though even reasonably large firms often have marginal products of 60%, or even 100%.

As long as we remain the world of aggregative growth theory, the average marginal product is of course equal to the marginal product, since marginal products are always equated. Moreover even if there is some transitory variation in the marginal product, the relevant number from the point of view of any investor, should the maximum and not the average: Capital should flow to where the returns are highest. The investments with returns of 60% or more should be the ones that guide investment, and not the 22%, and this ought to favor convergence.

That being said, there is nothing in what we have said that tells us whether 22% is lower than what we would have predicted based on an aggregative growth model that predicts convergence, or exactly right. Lucas (1990), in a well-known paper, suggests an approach to this question. He starts with the observation that according to the Penn World Tables (Heston, Summers and Aten (2002)), in 1990, output-per-worker in India at Purchasing Power Parity was 1/11th of what it was in the U.S. To obtain a productivity gap per effective use of labor, we need to adjust this ratio by the differences in education between the two countries. Based on the work of Krueger (1967), Lucas (1990) argues that "one American worker is equal to five Indian workers" in terms of human capital. In our case, since we are comparing productivity in 1990, and Krueger's estimates of human capital are from the late 1960s, we presumably adjust the correction factor. Between 1965 and 1990, years of schooling among those 25 years or older went from 1.90 years to 3.68 years in India and from 9.25 years to 12 years in the United States, i.e., from approximately 20% of the U.S. level, which fits with the 5:1 gap in productivity that Krueger suggested, to about 30%.[9]

[9] These numbers are based on Barro and Lee (2000). Another angle from which this can be looked at is that health

To show what this implies, Lucas starts with the assumption that *net* output is produced using a production function $Y = AL^{1-\alpha}K^{\alpha}$, where K is investment and L is the number of worker.[10]

From this, it follows that ouput per worker is $y = Ak^{\alpha}$, where k is investment per worker in equipment. Assuming that firms can borrow as much as they want at the rate r, profit maximization requires that $\alpha A k^{\alpha-1} = r$, from which it follows that

$$\frac{y_U}{y_I} = \left(\frac{r_I}{r_U}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{A_U}{A_I}\right)^{\frac{1}{1-\alpha}}. \tag{3}$$

If we assume that the only difference between the TFP levels in the two countries is due to the productivity per worker, the fact that Indian workers are only 30% as productive as the US workers and the share of capital is assumed to be 40% implies that:

$$\frac{A_U}{A_I} = (0.3)^{0.6} \approx 2. \tag{4}$$

With these parameters, the 11-fold difference between y_U and y_I would imply that $r_I = (3)^{5/3} r_U \approx 5r_U$. r is naturally thought of as the marginal product of capital. In other words, if we take 9% for the marginal product of capital in the U.S., this would imply a 45% rate for India

Lucas, at this point, did not even wait to look at the data: If the difference in the returns were indeed so large, all the capital would flow from the U.S. to India. Hence, Lucas argued, the rate of returns cannot possibly be that high in India. As we know, this is something of a leap of faith, since capital does not flow even when there are large differences in returns within the same country.

On the other hand, our estimates of the average marginal product is 22%. So Lucas was right in insisting that the actual rates of returns are much lower than what we would expect if the model were correct.

This is strictly only true if we estimate the marginal product from the data on output per worker; however if we calculate it directly from the capital-labor ratio, the problem shows up elsewhere. To see this, recall from equation 4 that assuming that workers are only 30% as productive is equivalent to assuming that TFP in India should be approximately 50% of what it is in the U.S. This, combined with the fact that, according to the Penn World Tables, the U.S. has 18 times more capital-per-worker than India implies that the marginal product of capital ought to be $\frac{1}{2}(18)^{0.6} = 2.8$ times higher in India, which tells us that the marginal product in India ought to be about 25%, which is probably close to what it is.

improved also during the period: Over a slightly different period, (1970-75 to 1995-2000), according to the Human Development Report (United Nations Development Program (2001)), life expectancy at birth went from 50.3 to 62.3 years in India and from 71.5 years to 76.5 years in the U.S., reducing the gap between India and the U.S. by about 40%.

[10] Lucas actually computes the ratio of output per effective unit of labor, which, with our parameters, is equal to $11 * \frac{3}{10} \approx 3$. Reassuringly, this is also the ratio that Lucas started with, albeit based on the average numbers for the 1965-1990 period rather than the 1990 numbers.

However if we now put in the numbers for capital-per-worker into the production function, the ratio of output per worker in the two countries turns out to be:

$$\frac{y_u}{y_I} = 2(\frac{k_U}{k_I})^{0.4} = 2 * (18)^{0.4} = 6.35. \tag{5}$$

In the data this ratio is 11:1. In other words, the problem is still there: Earlier when we used the capital-labor ratio implied by the low level of worker productivity it told us that the return on capital should be much higher than it is. On the other hand, when we use the actual capital-labor ratio, we see that the implied return on capital is quite reasonable, but the predicted worker productivity is much higher than it is in the data. Either way, it seems clear that we need to go beyond this model.

2.2 Investment Rates in Poor Countries

2.2.1 Is investment higher in poor countries?

Prima facie, it does not seem to be the case that investment rates are higher in poor countries. On the contrary, there is a robust positive correlation between investment rates in physical capital and income per capita, when both are expressed in terms of purchasing power parity. In fact, Levine and Renelt (1992) and Sala-I-Martin (1997) identified investment per capita as the only robust correlate of income. For example, Hsieh and Klenow (2003) estimate that in 1985, the correlation between PPP investment rate and PPP income per capita for the 115 countries present in the Penn World Tables was 0.60. The coefficients they estimate suggest that an increase in one log point in income per capita is associated with about a 5 percentage point higher PPP investment rate (the mean investment rate is 14.5%). The same positive correlation obtains with investment in plant and machinery. The relationship between investment rate and income per capita is much less strong when both of them are expressed in nominal terms rather than in PPP terms (Eaton and Kortum (2001); Restuccia and Urrutia (2001) and Hsieh and Klenow (2003)). The coefficient drops by a third when all investments are considered, and becomes insignificant when the measure of investment includes only plant and machinery. According to Hsieh and Klenow (2003), the fact that poor countries have a lower investment-to-GDP ratio, when expressed in PPP, is explained by the low relative price of consumption, relative to investment: While there is no correlation between investment prices and GDP, there is a strong positive correlation between consumption prices and GDP. It is not clear, however, that knowing this helps us explain why there is not *more* investment in poor countries. First, because the high rates that we found in some firms in developing countries and the lower, but still much higher than US, rates that we found on average are there despite the high price of capital goods. This, by itself, should encourage investment, unless income effects are unusually strong. Moreover, even if we measure everything in nominal terms, there is no strong negative correlation between investment and GDP.

There are, of course, examples of poor countries with large investment to GDP ratios. Young (1995) shows that a substantial fraction of the rapid growth of the East-Asian economies in the post-WWII period can be accounted for by rapid factor accumulation (including increase in the size of the labor force, factor reallocation, and high investment rates). In particular, according to the national accounts, between 1960 and 1985, the capital stock in Singapore, Korea, and Taiwan grew at more than 12% a year (in Hong Kong, it grew only at 7.7% a year). Between 1966 and 1999, the capital-output ratio has increased at an average rate of 3.4% a year in Korea, and 2.8% in Singapore. In Singapore, for example, the constant investment-to-GDP ratio increased from 10% in 1960 to 47% in 1984. In Singapore, Korea, and Taiwan, this increase in the stock of capital alone is responsible for about 1% out of the average yearly 3.4% to 4% of the "naive" Solow residual. Based on these results, Alwyn Young (Young (1995)) concluded that the East-Asian economies are perfect examples of transitional dynamics in the neo-classical model. However, in subsequent research, Hsieh (1999) questioned the validity of the national account data for investment for Singapore. He observes that if the capital-to-GDP ratio had grown at that speed, one would have observed a commensurate reduction in the rental price of capital. In practice, there was indeed a steady fall in the rental price of capital (both the interest rates and the relative price of capital fell) in Korea, Taiwan and Hong Kong. The drop is particularly large in Korea, where the national account statistics also suggest a large increase in the capital stock. However, in Singapore, there is no evidence that the rental rate declined over the period. If any thing, it seems to have *increased*.

As for investment in physical capital, there is no *prima facie* evidence that poor countries invest more in education. The data is poor and extremely partial, since it is difficult to estimate private expenditure on education. What we can measure easily, government expenditure on education as a fraction of GDP, however, is not higher in poor countries, though there is significant variation across countries. In 1996, according to the country level data disseminated by the World Bank "edstat" department, government investment on education was 4.8% in Africa, 4% in Asia, 4.1% in Latin America, 4.8% in North America and 5.6% in Europe. The correlation between the log of government expenditure on education as a fraction of GDP and GDP-per-capita is strong (in current prices): The coefficient of the log of GDP was 0.18 in 1990, and 0.08 in 1996, larger than the comparable estimate for rate of investment in physical capital.

As we noted earlier, the fact that teachers are relatively more expensive in developing countries may imply that true returns to education may be much lower than the Mincerian returns. Can this explain why there is not greater investment in education in poor countries? Within the neo-classical model, the answer is no: Banerjee (2004) shows that in the neo-classical world the same forces that raise the relative price of teachers in poor countries (or in countries with low education levels) also raise the wages paid to educated people, and on net the rate of return has to be higher rather than lower. And, in any case, it is

not true that public investment in education is higher when returns are higher: We found no correlation between government expenditure on education as a fraction of GDP and rate of returns to education (the coefficient of the rates of return to education on government expenditure in education in 1996 is -0.008, with a standard error of 0.013).

In summary, while there are isolated cases of high investment rates in relatively poor countries (Taiwan and Korea), this by no means seems to be a general phenomenon. We have already suggested one reason why this might be the case–it does not look like returns are especially high. It may also be that investment is not particularly responsive with respect to returns. This is the issue we turn to next.

2.2.2 Does investment respond to rates of return?

There is little doubt that people do take up many investment opportunities with high potential returns. Investment flowed into Bangalore when it became a hub for the software industry in India. When, in the 1990s, Tirupur, a smallish town in South India, became known in the U.S. as a good place to contract large orders of knitted garments, the industry in the city grew at more than 50% per year, due to substantial investments of both the local community (diversifying out of agriculture) and outsiders attracted to Tirupur (Banerjee and Munshi (2004)). Or, to take a last example from India, new hybrid seeds and fertilizers spread rapidly during the "green revolution", leading to very rapid yield growth (yields were multiplied by 3 in Karnataka and 2.5 in Punjab (Foster and Rosenzweig (1996))).

However, there are many instances where investments options with very high rates of returns do not seem to be taken advantage of. For example, Goldstein and Udry (1999) find that, despite the high rates of returns to growing pineapple compared to other crops, only 18% of the land is used for pineapple farming. Similarly, Duflo et al. (2003) find that only less than 15% of maize farmers in the area where they conducted field trials on the profitability of fertilizer report having used fertilizer in the previous season, despite estimated rates of return in excess of 100%.

From a more macro perspective, Bils and Klenow (2000) argue that the observed high correlation between educational attainment and subsequent growth observed in cross-sectional data (one year of additional schooling attainment is associated with 0.30 percent faster annual growth over the period 1960-1990) must be due, at least in part, to the fact that higher expected growth rates increase the returns to schooling, and therefore the demand for schooling. As we noted earlier, the correlation between education and subsequent growth (found in many studies, e.g., Barro (1991), Benhabib and Spiegel (1994), and Barro and Sala-I-Martin (1995)) appears to be too high to be entirely explained by the causal effect of transitional differences in human capital growth rates on growth rates. Bils and Klenow (2000) calibrate a simple neo-classical growth model, which requires that the impact of schooling on individual productivity has to be consistent with the average coefficient obtained from Mincer regressions. Their calibration

suggest that the high level of education in 1960 can only explain up to a third of the correlation between education and growth. Moreover, as we will discuss below, this correlation cannot be explained by high human capital externalities. They therefore calibrate an alternative model, where they construct the optimal schooling predicted by a country's expected economic growth. The calibration, once again, requires that the impact of education on human capital be consistent with the micro-estimates of the Mincerian returns, so that there remains a large fraction of the correlation between education and growth to explain. Higher expected growth induces more schooling by lowering the effective discount rate. They assume that a country's expected growth is a weighted average of its real *ex post* growth and the growth of the rest of the world. They estimate that, starting at 6.2 years of schooling, a 1 percent increase in growth induces 1.4 to 2.5 more years of schooling, depending on the values chosen for the parameters that are imposed. A 1 percentage point higher Mincerian return to schooling increases education by 1.1 to 1.9 years.

The aggregate data is thus consistent with a strong response of schooling to growth. However, it is also consistent with the presence of an omitted variable explaining both education and growth: In fact, Bils and Klenow acknowledge that their estimates suggests an elasticity of schooling demand to returns to schooling that is higher than what is implied by existing micro-studies (reviewed by Freeman (1986)). This problem cannot really be adequately addressed in the macroeconomic data, since there it is difficult to find a plausible instrument for growth, and the impact of expected growth on schooling must essentially be estimated as a residual impact (what remains to be explained from the correlation between growth and schooling after a plausible estimate for the impact of education on growth has been removed).

Foster and Rosenzweig, in a series of papers, use the green revolution in India as a source of partly exogenous increase in rate of returns to human capital to estimate the impact of expected growth and increases in returns to education on schooling and, more generally, investment in human capital. Foster and Rosenzweig (1996) find that returns to education increased faster in regions where the green revolution induced faster technological change: Their estimates imply that in 1971, before the start of the green revolution, the profits in households where the head had completed primary education were 11% higher than the profits in households were he had not. By 1982, the profits were 46% higher for districts where the growth rate was one standard deviation above average. They then turn to estimating whether educational choices were also sensitive to the higher yield growth. After instrumenting for yield growth, they find that the impact of technological change on education is indeed substantial: In areas with recent growth in yields of one standard deviation above the mean, the enrollment rates of children from farm households are an additional 16 percentage points (53%) higher, compared to average-growth areas. Foster and Rosenzweig (2000) find that technological growth also affected the provision of schools, benefiting landless households. However, on balance, technological growth seems to lead to lower educational investment by

landless households, perhaps because returns to education increase less for them (since they are engaged in more menial tasks) and because the fact that the withdrawal of children of landed households from the labor market increases children's wages, and thus the opportunity cost of school attendance.

Foster and Rosenzweig (1999) consider another measure of investment in children's human capital, namely child survival. They argue that technological growth in the village increases the returns from investing in boys' health, while technological growth outside the village, but in the potential "marriage market", increases the returns to investing in girls (because better educated and healthier women will fetch a higher prices in regions with higher technological progress). Their results indeed suggest that the gap in boys/girls mortality rates increases with technological change in the village, but decreases with technological change in the labor market.

Other evidence that girls' survival is affected by the expected returns to having girls include Rosenzweig and Schultz (1982), who show that the boys/girls mortality gap is negatively correlated to women's wage, and Qian (2003), who uses the liberalization of tea prices in China as a natural experiment in female productivity. She shows that, in regions suitable to tea production, the ratio of boys to girls diminished considerably after tea production and tea prices were liberalized. Since tea is picked by women, this is evidence that higher female productivity encourage parents to invest more in their girls. In contrast, in regions suitable for orchard production (for which male have an advantage) the ratio of boys to girls increased during the period.

While these facts taken together do suggest that individuals do respond to returns when making human capital investment decisions, there are possible alternative explanations for these facts. The results from Rosenzweig and Schultz (1982) and Qian (2003) cannot easily be distinguished from a women's bargaining power effect: If mothers tend to prefer girls, and their bargaining power increases as a result of the increase of their productivity, then the outcomes will improve for girls, even if households' decisions do not respond to returns. The results in Foster and Rosenzweig (1996, 2000) could in part be attributed to wealth effects (expected growth makes the households richer, and if education has any consumption value, one would expect growth to respond to it), although Foster and Rosenzweig (1996) estimate the wealth effect directly, and argue that it is not important. But it remains possible that the instrumented expected increase in yield captures real increases in expected wealth better than any other measure (they show that land prices do adjust to the future expected yield increases, for example). Moreover, there is also direct evidence that investment in human capital does not always respond to returns: Munshi and Rosenzweig (2004) show that the rapid increase in the returns to English education in India in the 1990s (the returns increased from 15% to 24% in 10 years for boys, and 0% to 27% for girls) led to a convergence in the choice of English as a medium of instruction between the low and high castes amongst girls, but not amongst boys: Boys from the lower castes seem so far not to have taken

full advantage of the new opportunities offered by English medium education.

Another angle for approaching this question is the sensitivity of human capital investment to the direct or indirect costs of these investments. Several recent studies do suggest that the elasticity of school participation with respect to user fees is high: Kremer, Moulin and Namunyu (2003) conducted a randomized trial in rural Kenya in which an NGO provided uniforms, textbooks, and classroom construction to seven schools randomly selected from a pool of 14 schools. Dropouts fell considerably in the schools that received the program, relative to the other schools (after five years, pupils initially enrolled in the treatment schools had completed 15% more schooling than those enrolled in the comparison schools). They argue that the financial benefits of the free uniforms were the main reason for this increase in participation. Several programs go beyond reducing the school fees to actually pay for attendance. The PROGRESA program in Mexico provided grants to poor families, conditional on continued school participation and participation in health care. The program was initially launched as a randomized experiment, with 506 communities randomly assigned to either the treatment or control group. Schultz (2004) finds a 3.4% increase in enrollment in all children. The largest increase was in the transition between primary and secondary school, especially for girls. Gertler and Boyce (2002) report a similar effect on health. In this case as well, it is difficult to distinguish the pure price effect from the income effect.[11] School meals, which is another way to pay children to attend school, have been shown to be associated with increased school participation in several observational studies (Jacoby (2002); Long (1991); Powell, Grantham-McGregor and Elston (1983); Powell, Walker, Chang and Grantham-McGregor (1998) and Dreze and Kingdon (2001)) and one experimental trial conducted among pre-school children in Kenya (Vermeersch (2002)). The available evidence, therefore, points toward a robust elasticity of schooling decisions with respect to the cost of schooling.

While this could be indicative of households being extremely sensitive to net returns, the magnitude of these effects are hard to reconcile with this explanation. For example, using an estimate of 7% Mincerian returns per year of education, Miguel and Kremer (2004) estimate that the benefit of one year of primary schooling is in excess of $200 over the lifetime of a child. Yet, the provision of a uniform valued at $6 induced an average increase of 0.5 years in the time a child spent in school (time spent in schools increased from 4.8 years in the comparison schools on average, to 5.3 years in the treatment schools). To be consistent with a model where the only reason where the provision of uniforms increase school attendance is the increase in the rate of returns that it leads to, these numbers would mean that a large fraction of children (or their parents) were exactly indifferent between attending school or not, before the uniform is provided.

While this is certainly possible, other evidence suggests that human capital investment does not always

[11] Moreover, there could be a bargaining power effect, since the grants were distributed through women.

respond to rates of returns. For example, the take-up of the de-worming program studied by Miguel and Kremer (2004) was only 57%, despite the fact that the program was *free*, and that the only investment required was to sign an informed consent form (and some disutility for the child). Further, when a nominal fee was introduced in a randomly selected set of schools in the year after the initial experiment, the take-up fell by 80%, relative to free treatment (Kremer and Miguel (2003)). While this could be due to the fact that the private benefits are perceived to be low by the parents, it is worth noting that the hike in user fees happened after one year of free treatment, so that parents would have had time to observe the change in the child's health and attendance at school. Moreover, Kremer and Miguel (2003) also observe that, as long as the price was positive, there was no impact from the actual price on the take-up of the drug. This strong non-linearity between a price of zero and any positive price (which is also consistent with the evidence from school uniforms) appears to be inconsistent with an explanation of their findings in terms of rates of returns.

To sum up, the evidence suggests that, while investment seems to respond in part to the cost and the benefits of these investments, it appears to do so in ways that suggest that it does not only respond to returns as we are measuring them.

2.2.3 Taking Stock: Investment Rates

Investment rates, both in physical and human capital are typically no higher in poorer countries than in rich countries. If we are willing to accept that the average marginal product is the one that guides investment, this is perhaps not a surprise, especially given that the link between investment and rates of return is also not particularly strong.

3 Understanding Rates of Return and Investment Rates in Poor Countries: Aggregative Approaches

For Lucas, the inability to fit the cross-country differences into an aggregative growth model was direct evidence in favor of abandoning the assumption of equal TFP: Allowing the TFP level to be lower in India than in the US will increase the difference in output per worker between the two countries, for any given difference in factor endowments. This is also the message of the more recent literature on development accounting (Klenow and Rodriguez-Clare (1997) , Caselli (2004)), who demonstrate that it is impossible to explain even half of the cross-country variation in output per worker based on variation in the stocks of capital and human capital, even after making adjustments for the quality of these inputs, and other possible sources of mismeasurement.

Allowing TFP differences across countries can also explain why there poor countries do not invest more and ultimately why there is no growth convergence: Once we assume fixed productivity differences, the steady states in different countries will be different and there is no presumption that poorer countries should grow faster.

For rest of this section, we will therefore proceed under the assumption that it is possible to resolve the "macro puzzles" by introducing cross-country differences in TFP. We focus on the so-called new growth approaches, which are theories within the aggregative growth framework that aim to explain persistent cross-country TFP differences, recognizing that these "new" growth theories, like "old" growth theories, make no attempt to deal with the obvious problems with the aggregate production function.

3.1 Access to Technology and the Productivity Gap

The dominant answer, within growth theory, of why TFP should be lower in poorer countries comes down to technology. There is a now a large literature—due to Aghion and Howitt (1992), Grossman and Helpman (1991) and others—that, emphasizes technological differences as the source of this TFP gap. It is easy to think of reasons why there may be a persistent technology gap between rich and poor countries. Essentially, it is too costly for the poor country to jump to the technological frontier because the frontier technologies belong to firms in the rich countries (who are the ones who have the biggest stake in keeping the technological frontier moving) and they charge monopolistic prices for access to these technologies.[12] Moreover, there is the issue of appropriate technology: The latest technology may not be suitable for use in a country with little human capital[13] or poor infrastructure.

By itself, this explanation focuses on investment in technology and cannot directly account for the lack of investment in human capital in LDCs or why the returns there often seem so low. However, if there is strong complementarity between human capital investment and investments in new technology,[14] then the slow growth of TFP could explain the relative absence of investment in human capital in LDCs, assuming that we accept the rather mixed evidence, reviewed above, on the responsiveness of investment in human capital to the expected returns.

If the productivity gap between the U.S. and India has to be fully accounted for by technological differences in an aggregative model (i.e., if we rule out any differences in the interest rates), then TFP in the U.S. would have to be about twice that in India. How plausible is a TFP gap of 1:2 in a world of efficiently functioning markets? One way to look at this is to observe that U.S. TFP growth rates

[12] The dominance of rich countries in the latest technologies is reinforced by the fact that the rich countries may have an actual advantage in R&D, because of their larger market size or their superior human capital endowments.

[13] For example, as in Acemoglu and Zilibotti (2001) or Howitt and Mayer (2002)

[14] As suggested, for example, in the work by Foster and Rosenzweig (1995) on the green revolution, which we discussed above.

seem to be on the order of 1-1.5% a year. Even at 1.5%, TFP takes about 45 years to go up by 200%.[15] Therefore in 2000, Indians would have been using machines discarded by the U.S. in the 1950s.

This is also clearly very far from being true of the better Indian firms in most sectors. The McKinsey Global Institute's (McKinsey Global Institute (2001)) recent report on India, reports on a set of studies of the main sources of inefficiency in a range of industries in India in 1999, including apparel, dairy processing, automotive assembly, wheat milling, banking, steel, retail, etc. In a number of these cases (dairy processing, steel, software) they explicitly say that the better firms were using more or less the global best practice technologies wherever they were economically viable. The latest (or if not the latest, the relatively recent) technologies were thus both available in India and profitable (at least for some firms).

However, most firms do not make use of these technologies. And, according to the same McKinsey report, it is not because these technologies are not economically viable in this sector: The report on the apparel industry tells us that in the apparel industry:

> "Although machines such as the spreading machine provide major benefits to the production process and are viable even at current labor costs, they are extremely rare in domestic (i.e., non-exporting) factories." (McKinsey Global Institute (2001))

Despite this, technological backwardness is not one of the main sources of inefficiency that is highlighted in their report on the apparel industry. They focus, instead, on the fact that the scale of production is frequently too small, and in particular, on the fact that the median producer is a tailor who makes made-to-measure clothes at a very small scale, rather than a firm that mass produces clothes. TFP is low, not because the tailors are using the wrong technology given their size, but because tailoring firms are too small to benefit from the best technologies and therefore should not exist.

Reports from a number of other industries show a similar pattern. Certain specific types of technological backwardness is mentioned as a source of inefficiency in both the dairy processing industry and the telecommunications industry, but in both cases it is argued that while all firms should find it profitable to upgrade along these dimensions (McKinsey Global Institute (2001)), only a few of them do.

In these two cases, however, there is also a reference to the gains (in terms of productive efficiency) from what the report calls "non-viable automation". This is automation that would raise labor productivity but lower profits. One reason why automation may be non-viable in this sense is that the technology may be under patent and therefore expensive, along the lines suggested by Aghion and Howitt (1992), Grossman and Helpman (1991) and others, or it may demand skills that the country does not have. On

[15] The effective rate of technological improvement will be larger, for example, if new technology needs to be embodied in machines and machines are more expensive (or savings rates are lower) in the poorer country (see Jovanovic and Rob (1997)).

the other hand, it could also be something entirely neo-classical: Labor-saving devices are less useful in labor-abundant countries. Since we have no way of determining why the technology is non-viable, we looked at the total labor productivity gain promised by this category of innovations. In both the dairy processing industry and the telecom case, this number is 15% or less, and in the automotive industry it is no larger (McKinsey Global Institute (2001)). This is clearly nowhere near being large enough to explain the entire TFP gap.

On other hand, it is clearly true that there are many firms that, for some reason, have opted not to adopt the best practice despite the fact that others within the same economy find it profitable to do so and, at least according to McKinsey, they too would benefit from moving in this direction. In other words, while there is a technology gap, it is largely a within-country phenomenon and not, as the models of technology production and adoption imply, a problem at the level of the country.[16]

3.2 Human capital externalities

Another reason why there may be persistent TFP differences across countries, is that there are aggregate increasing returns. As emphasized in the introduction, for this to be true it is not enough to have firm level increasing returns. We need externalities across firms, or more generally across investors, which may arise, for example, because there are human capital externalities: It has been argued that human capital is not just valuable to those who own and use it, but also to others.

Externalities in human capital would tend to limit the extent of diminishing returns with respect to human capital in the production function, keeping as given the share of human cpital in total production. This would tend to raise productivity in rich countries (who have a lot of human capital) and slow down convergence.

Externalities could also explain a puzzle we did not discuss until now, pointed out by Acemoglu and Angrist (2001) and Bils and Klenow (2000): The high correlation between human capital and income that is observed in the cross-country data (e.g., Mankiw et al. (1992)) is hard to reconcile with the micro evidence we have reviewed earlier, which suggested relatively low returns to education. To see this, note that the difference in average schooling between the top and bottom deciles of the world education distribution in 1985 was less than 8 years. With a Mincerian returns to schooling of about 10%, the top decile countries should thus produce about twice as much per worker as the countries in the bottom decile. In fact, the output-per-worker gap is about 15. One possibility is that the Mincerian rate of

[16] It is true that machines of different vintages, and therefore different productivities, can co-exist even when markets work perfectly, as long as machines are long-lived (Bardhan (1969)). However the technological gap between the tailor and the garment factory (or the milk-man and the fully-automatized dairy) seem too large and too stable to be just a transitional problem.

return understates the true rate of returns to education, because it does not take into account positive externalities generated by educated workers. More specifically, the human capital externalities on the order of 20-25% (more than twice the private return) would be necessary to explain the cross-country relationship between education and income, which sounds implausible.

Early evidence (e.g., Rauch (1993)) suggested that externalities were positive, but not of that order of magnitude. Using variation in education across U.S. cities, Rauch (1993) estimated that the human capital externalities may be on the order of 3% to 5%. Moreover, even this evidence is to be taken with caution, since cities where workers are more educated vary in many other respects. Using variation in average education generated by the passage of compulsory schooling laws, Acemoglu and Angrist (2001) find no evidence of average education on individual wages, after controlling for individual education.

In Indonesia, Duflo (2003) actually finds evidence that those who invest in their education may inflict negative pecuniary externalities on others. She studies the impact of an education policy change that differentially affected different cohorts and different regions of Indonesia. Between 1973 and 1979, oil proceeds were used to construct over 61,000 primary schools throughout the country. Duflo (2001) shows that the program resulted in an increase of 0.3 years of education for the cohorts exposed to the program. Duflo (2003) takes advantage of the fact that individuals that were 12 or older when the program started did not benefit from the program, but worked in the same labor markets as those who did. As the newly educated workers entered the labor force, starting in the 1980s, the fraction of educated workers in the labor force increased. Since migration flows in Indonesia remained relatively modest, the increase in the fraction of workers with primary education between 1986 and 1999 was faster in regions which received more INPRES schools. Using the interaction of year and region as instruments for the fraction of educated workers, she estimates that an increase of 10 percentage points in the fraction of educated workers in the labor force resulted in a *decrease* in the wages of the older workers (both educated and uneducated) by 4% to 10%. This suggests that, on balance, there are strongly diminishing aggregate returns at the local level: Any positive externality is more than compensated by these declining returns. The Mincerian returns could then actually overestimate the aggregate returns of increasing education, because by comparing individuals within a labor market, they do not take into account the diminishing returns that affect everybody in the labor market.

To summarize, the available evidence does not suggest that there are strongly increasing returns to human capital. It appears that if human capital externalities are important they must take a very different form. One possibility is that they play a role in the intergenerational transmission of learning: For example, it is possible that parents or teachers do not fully internalize the benefits that their investment in human capital confer on the next generation of scholars. However as the figures above makes clear, in order to fit the data the extent of this miscalculation has to be substantial.

3.3 Coordination failure

Another source of lower aggregate productivity is the possibility of coordination failures, which reduces aggregate productivity through a demand effect. There is a long line of work, starting with Rosenstein-Rodan (1943), that has emphasized the role of coordination failure in explaining why certain countries successfully industrialize, while others remain poor and non-industrialized. Murphy, Shleifer and Vishny (1989) explore models where industrialization in a sector creates demand for the products of another sector (through higher wages for the workers), and which leads to multiple equilibria. A coordinated "big push", where all industries start together, can place the country on a permanently higher level of investment and income. Developing countries may have low investments and low returns to capital because such a "big push" has not happened. A large literature explores different forms of strategic complementarities. Since the argument involves an entire economy's coordination, it is difficult to use micro-evidence to provide much direct evidence about these aggregate externalities.[17] However, while these theories certainly have some relevance, the fact countries trade will tend to substantially mitigate the effect of local demand. It is therefore not clear that aggregate demand effects can be so powerful as to generate the necessary gap in TFP between, say, India and the US.

Another possible mechanism is suggested by Acemoglu and Zilibotti (1997). They argue that when one firm invests, others benefit, because each firm is subject to independent shocks and increasing the number of firms expands the ability of an individual firm to diversify its risk. When there are few firms around, risk diversification opportunities are limited and risk averse investors limit their investment.

The India-US comparison is perhaps the worst example one could pick for applying this theory, since between the 1920s and the 1940s the Indian stock market was comparable to that in many OECD countries in terms of number of listed firms and volume of trade. However lack of financial development is clearly a serious problem for many developing countries. However, within an aggregative model the lack of financial development can at best explain why all the firms underinvest. As emphasized earlier, the more important sources of inefficiency seem to come from the fact that some firms underinvest much more than others, and in particular that some firms adopt the latest technologies, while others do not.

3.4 Taking Stock

While the evidence is somewhat impressionistic, it seems unlikely that the aggregative theories discussed above, can explain the entire TFP gap. Of course, if we were prepared to give up the idea that the entire problem comes from a lower aggregate productivity, for example by accepting that the marginal product is lower in India, the problem of fitting the data would be easier. For example, if the TFP gap

[17] Below, we will review the evidence on more local externalities.

were 1.5 higher in the US (on top of what is predicted by the difference in the productivity of labor), the fact that the U.S. has 18 times more capital-per-worker would imply that output-per-worker would be $(1.5)(2)(18)^{0.4} = 9.5$ times higher in the U.S., and the marginal product of capital would be $\frac{(18)^{0.6}}{2(1.5)} = 1.9$ times higher in India. These are both clearly in the ballpark, although the output gap between the U.S. and India predicted by this model is still too low (the output gap is about 11:1 in the data) and the ratio of the marginal product of capital between India and the U.S., which was too high in a model with identical TFP, is now too low (the ratio in the data is about 2.5).

It is worth noting that in order to get closer to 11:1 ratio in the data, the TFP ratio would need to be higher than 1.5, which is perhaps already too big. Moreover, this would further reduce the predicted ratio between the marginal product of capital in India and in the U.S., which was already too low when the TFP gap was 1.5. In other words, we are facing a new problem: Given the existing capital stock, if a difference in TFP was the reason why the output per worker is so low in India, the marginal product of capital should be even lower in India than what it is. Indeed, there is no way to adjust the TFP ratio to improve the fit along both dimensions—we can increase the gap in output-per-worker by raising the TFP ratio, but only at the cost of making the ratio of marginal product even smaller. The problem is quite basic: With a Cobb-Douglas production function, the average product of capital is proportional to its marginal product. But then output-per-worker must be proportional to the product of the marginal product of capital and capital-per-worker. If the marginal product in India is 2.5 times that of the U.S., but capital-per-worker is 18 times greater in the U.S., output-per-worker has to be $\frac{18}{2.5} = 7.2$ times larger in the U.S. and not 11 times larger, *irrespective of what we assume about the ratio of TFP in the two countries.* In other words, the only way we can hope to really fit what we see in the data is by abandoning the standard Cobb-Douglas formulation. This is useful to keep in mind when, in later sections, we discuss ways to improve the fit between the theory and the data.

To sum up, Lucas' question about why capital does not flow from the U.S. to India was, in some sense, where it all started, but from the vantage point of what we know today, this is in some ways the lesser problem. We know now that there are differences in the marginal product of capital within the same economy that dwarf the gap that Lucas calculated from comparison of India and the U.S., and found so implausibly large that he set out to rewrite all of growth theory. The harder question is why capital flows do not eliminate these differences.

Lucas' resolution of the puzzle, was to give up the key neo-classical postulate of equal TFP across countries. Based on the McKinsey report, this seems to be the obvious step, but the problem is less that people in developing countries do not find it profitable to adopt the latest (and best) technologies and more that many firms do not adopt technologies that are available and would be profitable if adopted. The key question, once again, is why the market allows this to be the case.

The premise of the aggregative approach to growth was that markets function well enough within countries that we can largely ignore the fact that there is inefficiency and unequal access to resources within an economy when we are interested in dynamics at the country level. The evidence suggests that this is not true: The cross-country differences in marginal products or technology that we want to explain are of the same order of magnitude as the differences we observe within each economy. A theory of cross-country differences has to based on an understanding and an acknowledgment of the reasons why rates of returns vary so much within each country. This is what we turn to next: In section 4, we first review the various reasons that have been proposed. In section 5, we will then calibrate their impact to evaluate whether they can form the basis of an explanation for the puzzles we observed.

4 Understanding Rates of Return and Investment Rates in Poor Countries: Non- Aggregative Approaches

In this section, we review various possible reasons why individuals do not always make the best possible use of resources available to them.

4.1 Government Failure

One reason why firms may not choose the latest technologies or make the right investments is because they do not have the proper incentives to do so. A line of work has developed the hypothesis that governments are largely responsible for this situation, either by not protecting investors well enough or by protecting some of them excessively. The firms that are ill-protected underinvest and have high marginal returns, while the over-protected firms overinvest and show low marginal returns. The net effect on investment may be negative, because even those who are currently favored may fear a future falling out and a corresponding loss of protection. Overall productivity may also go down, since the right people may not always end up in the right business, since connections rather than skills will dominate the choice of professions.[18]

One approach to investigating this hypothesis has been to try to document variations in the quality of institutions, and to try to evaluate their impact. La Porta, de Silanes, Shleifer and Vishny (1998) document important variations in the degree to which the law protects investors (creditors and shareholders) across countries, part of which seem to be explained by the origin of these countries' legal codes (the French civil law has much less legal protection for investors than Anglo-Saxon common law). Djankov, Porta, de Silanes and Shleifer (2002) document wide variation in the ability of someone to start a new

[18]See Murphy and Vishny (1995).

firm in 85 countries. They argue that the costs of entry are high in most countries (on average, they sum up to 47% of a country's GDP per capita), and can be very high indeed: While it take 3 procedures and 3 days to obtain the permit to start company in New Zealand, it takes 19 procedures, 149 business days and 111.5 percent of GDP per capita in Mozambique. The procedure is shorter, and generally less expensive in terms of GDP per capita, in rich countries than in poor or middle-income countries. Djankov, Porta, de Silanes and Shleifer (2003) document the time it takes in court to evict a tenant or collect a bounced check, as well as the degree of formalism of the legal procedures. They find, once again, wide variation: In particular, these procedures take a much shorter time in countries with common law legal origins. Similarly, many studies argue that, in cross-country regressions, there is a strong association between aggregate investment and measures of bad institutions or corruption (e.g., Knack and Keefer (1995), Mauro (1995), Svensson (1998)).

These papers also argue that low investor protections, legal barriers to entry, and long legal procedures have implications for welfare and efficiency. There are indeed suggestive associations in the data (for example, ownership is more concentrated when investor protection is worst), but there is always the possibility that the correlation between the quality of the institutions and the real outcomes they consider is due to a third factor. Acemoglu, Johnson and Robinson (2001) try to address this issue by finding exogenous variations in the quality of institutions. They argue that there is a persistence of institutions, so that countries which accessed independence with extractive institutions (e.g., Congo) have tended to keep these bad institutions. They then argue that colonial powers were more likely to set up extractive institutions, with an unrestrained executive power, in places where they did not intend to settle. Finally, they were less likely to settle in places where the environment was hostile: In particular, the mortality of early settlers predicted the number of people of European descent who settled in these countries, the quality of institutions at the turn of the 20th century, and the quality of institutions in 1995 (measured as the risk of expropriation perceived by investors). In turn, it also is associated with lower GDP in 1995. The authors then use early settler mortality as an instrument for institutions in a regression of the impact of institutions on inequality, and find a strong positive coefficient.

This evidence suggests that governments matter, and that bad governments will lower returns and discourage new investments. There is a literature that tries to investigate the exact mechanisms through which the government affects the allocation of resources. One version of the story blames excessive intervention, while another talks about the lack of appropriate regulations. We now discuss these two explanations in turn, and try to assess how far they can help us fit the evidence.

4.1.1 Excessive Intervention

There is a line of work, following Parente and Prescott (1994, 2000), which argues that the productivity gap results from the way the heavy hand of the government operates. The government makes rules that discourage entry and innovation and protects the inept, and thereby slows the economy's progress towards the ideal state where only the most productive firms survive. The regulation may lead the economy to have too few firms, leaving inefficient incumbents to run the firms (see Caselli and Gennaioli (2004) and other references in this study), or too many firms, when regulation discourages consolidation by treating small firms and larger firms differently.

There is clearly something to this vision. Gelos and Werner (2002) show that financial de-regulation in Mexico (which started in 1988 and eliminated the interest rate ceiling, high reserve requirements which channelled 72% of commercial bank lending to the government, and priority lending) increased the ability of small firms to access the credit market, and reduced the excess cash flow sensitivity of investment *for small firms only.* Until recently in India, a large number of sectors were reserved for firms below a certain size (the small-scale sector) and/or firms in the cooperative sector. Small firms also benefited from tax exemptions and priority sector credits. This clearly limited the ability to take advantage of economies of scale and restricted the market share of the most efficient players.

Nonetheless, this is probably only a part of the story. As we noted in the context of the discussion of Banerjee and Duflo (2004), even medium-sized firms that were well above the cut-off for being included in the small-scale sector seem to be operating well below their optimal scale. In other words, notwithstanding the politically protected presence of the small-scale firms that is presumably driving down profits in the sector, these medium-sized firms were clearly still at the point where further investment would be extremely profitable. There has to be something other than a policy-induced lack of profitability that was holding them back.

The same point is made in a different style in the paper by Banerjee and Munshi (2004), mentioned above. This paper studies investment and productivity differences among firms in the knitted-garment industry in Tirupur, India. The firms owned by the Gounders, tend to be much larger than the firms owned by all other participants in the industry: The gap among firms that had just started is on the order of three to one. Yet these Gounder firms produce much less per unit of capital, and Gounder firms that have been in business for more than five years actually produce less *in absolute terms* than the smaller firms of the same vintage owned by non-Gounders. In other words, it is the bigger firms that are less productive, in an environment where the government discriminates, if at all, in favor of the smaller firms.

To sum up, while there are certainly instances of excessive intervention, it seems that there are many inefficiencies that cannot be blamed on the government.

4.1.2 Lack of Appropriate Regulations: Property Rights and Legal Enforcement

Effective rates of return and investment rates can be low because the responsibilities and/or the benefits of the investments are shared, or the investors are worried about being expropriated: The investor is therefore not capturing the full marginal returns of its investment. Imperfect property rights will thus lead to low investments. Poorly enforced property rights also make it difficult to provide collateral, which exacerbates the problems of the credit market. For example, the study of the Mexican financial deregulation discussed above (Gelos and Werner (2002)) showed that after the deregulation, small firms' access to credit became more linked to the value of the real estate assets they could use as collateral: The role of the government does not end with not interfering, it may also be to provide secure property rights.

In addition to the macro-economic evidence mentioned above, there is some micro-economic evidence that property rights matter for investment, although the findings are more mixed. Goldstein and Udry (2002) show that, in Ghana, individuals are less likely to leave their land fallow (which is an investment in long run land productivity) if they do not hold a position of power within the family of the village hierarchy which ensures that their land is not taken away from them when it is fallow. However, Besley (1995) finds that, also in Ghana, investment (tree planting) is not significantly larger when individuals have more secure rights to their land. Johnson and Woodruff (2002) find that, in five post-Soviet countries, firms that are run by entrepreneurs who perceive that their property rights are more secure invest more than those who do not. The effect is as strong for firms who rely mostly on internal finances as for those who have access to external finance. Entrepreneurs who believe that they have strong property rights invest 56% of their profits in their firms (against 32% for those who do not). Do and Iyer (2003) find that a land reform which gave farmers the right to sell, transfer or inherit their land usage rights also increased agricultural investment, in particular the planting of multi-year crops (such as coffee).

Even when property rights themselves are legally well defined and protected, there are institutions which reduce the private incentives to invest. Sharecropping is one environment where both the landlord and the tenants have low incentive to invest in the inputs that they are responsible for providing (Eswaran and Kotwal (1985)). Binswanger and Rosenzweig (1986) and Shaban (1987) both show that, controlling for farmer's fixed effect (that is, comparing the productivity of owner-cultivated and farmed land for farmers who cultivate both their own land and that of others) and for land characteristics, productivity is 30% lower in sharecropped plots. Shaban (1987) shows that all the inputs are lower on sharecropped land, including short-term investments (fertilizer and seeds). He also finds systematic differences in land quality (owner-cultivated has a higher price per hectare), which could in part reflect long-term investment. Banerjee, Gertler and Ghatak (2002) study a tenancy reform which increased the tenants' bargaining power and security of tenure. They found that the land reform resulted in a substantial

increase in the productivity of the land (62%). Since the reform took place at the same time as the green revolution, this increase in productivity is probably in part due to an increased willingness to switch to the new seeds after the registration program.[19]

The example of sharecropping suggests that bad governments are not the only cause for the emergence of bad institutions. If sharecropping is inefficient, why does it arise? In particular, why do the landlord and the tenant not agree on a fixed rent, which will ensure that the tenant is the full beneficiary of his effort at the margin? Explanations of the persistence of sharecropping involve risk aversion (Stiglitz (1974)) or limited liability (Banerjee et al. (2002)). This suggests that while the proximate explanation for inefficient investment may well be based in a specific institution, the more basic cause may be lying elsewhere, in the way various asset markets function. This is what we turn to next.

4.2 The Role of Credit Constraints

• Why would credit markets function poorly in poor countries?

The fact that the capital market does not function well in poor countries is a result of a number of factors. First, information systems, including property records, are often underdeveloped, making it hard to enforce contracts. This, in turn, partly reflects the fact that people may not know how to read or write and partly the fact that there has not been enough institutional investment.[20] Second, the fact that potential borrowers are poor and under extreme economic pressure, might make them all too willing to try to cheat the lender. Third, there are political pressures to protect borrowers from lenders in most LDCs.

• Consequence of poorly functioning credit market.

Given the problems in enforcing the credit contract, what a lender will be prepared to offer a particular borrower will depend on the quality of the borrower's collateral, his reputation in the market, the ease of keeping an eye on him and a host of other characteristics of the borrower. This has the obvious implication that two firms facing the exact same technological options may end up choosing very different methods of production. In particular, one person may start a large or more technologically advanced firm because he has money and another may start a small and backward one because he does not.[21] As a result, neither

[19] This interpretation is reinforced by the fact that their estimates are higher than Shaban's and that of a study by Laffont and Matoussi (1995) who use data from Tunisia to show that a shift from sharecropping to fixed-rent tenancy or owner cultivation raised output by 33 percent, and moving from a short-term tenancy contract to a longer-term contract increased output by 27.5 percent.

[20] For example, Djankov et al. (2002) document the time it takes the recover a bounced check across countries. It takes longer in poorer countries.

[21] Aghion, Howitt and Mayer-Foulkes (2004) argue that credit constraints may also be important in explaining the cross-country differences in the adoption of new technologies.

interest rates, nor TFP, nor the marginal product need be equalized across borrowers.

This would also explain why investment responds so unpredictably to returns: Sometimes the opportunities become available when there is large group of people who are looking to invest and have the wherewithal to do it. At other times, the returns may be there but most of those who have money may be heavily involved in promoting something else.

A second set of implications of imperfect contracting in the credit market is that the supply curve of capital to the individual borrower slopes up—a borrower who is more leveraged will need more monitoring and the lender will charge him more to do the extra monitoring. And eventually, the extra monitoring may be too costly to be worth it, and the borrower will face an absolute limit on how much he can borrow.

An immediate consequence of an upward-sloping supply curve is that the marginal product of capital will be higher than what the borrower pays the lender. Indeed, the gap between the two may quite substantial, since the fact that borrowers are constrained in borrowing also implies that the lenders are constrained in how much they can lend at rewarding rates. This drives the interests rates down, as lenders compete for the best borrowers. Moreover, since the rates the lenders charge include the cost of the monitoring that they have to do, the rates the lenders charge could be much higher than the opportunity cost of capital. In the case of a financial intermediary, such as a bank, this implies that the rates they charge their borrowers may be much higher than the rates they pay their depositors.

This implies, for example, that the American investor who gets 9% on his stock market investment could not just put the money in a bank in India and earn the 22% average marginal product. Indeed, he may not earn much more than 9% if he were to put it in an Indian bank. However, he could set up a business in India and earn those returns, and presumably if enough people did that, the returns would be equalized; below we will try to say something about why this does not happen.

It also implies that the incentive to save may be low in countries where the marginal product is high, except for those who are planning to invest directly. This might help to explain the low equilibrium investment rate, though it is theoretically possible that the negative effect on the savers would be swamped by the positive effect on investors if the fraction of investors is large enough.

• Evidence

We have already mentioned some evidence from South Asia showing that the interest rate varies enormously across borrowers within the same local capital market and that the extent of variation is too large to be explained by the observed differences in default rates. Banerjee (2003) lists a number of studies that make it clear that this is also true in developing countries outside South Asia. This is suggestive, albeit indirect, evidence of credit constraints.

If the marginal product of capital in the firm is greater than the market interest rate, credit constraints naturally mean that a firm would want to borrow more than what is available. It is, however, not clear how

one should go about estimating the marginal product of capital. The most obvious approach, which relies on using shocks to the market supply curve of capital to estimate the demand curve, is only valid under the assumption that the supply is always equal to demand, i.e., if the firm is never credit constrained.

The literature has therefore taken a less direct route: The idea is to study the effects of access to what are taken to be close substitutes for credit—current cash flow, parental wealth, community wealth—on investment. If there are no credit constraints, greater access to a substitute for credit would be irrelevant for the investment decision. While this literature has typically found that these credit substitutes do affect investment,[22] suggesting that firms are indeed credit constrained, the interpretation of this evidence is not uncontroversial. The problem is that access to these other resources is unlikely to be entirely uncorrelated with other characteristics of the firm (such as productivity) that may influence how much it wants to invest. To take an obvious example, a shock to cash-flow potentially contains information about the firm's future performance.

The estimation of the effects of credit constraints on farmers is significantly more straightforward since variation in the weather provides a powerful source of exogenous short-term variation in cash flow. Rosenzweig and Wolpin (1993) use this strategy to study the effect of credit constraints on investment in bullocks in rural India.

The paper by Banerjee and Duflo (2004) that we discussed above makes use of an exogenous policy change that affected the flow of directed credit to an identifiable subset of firms in India. Since the credit was subsidized, an increase in sales and investment as a response to the increase in funds available needs to mean that firms are credit constrained, since it may have decreased the marginal cost of capital faced by the firm. However, they argue that if a firm is not credit constrained then an increase in the supply of subsidized directed credit to the firm must lead it to substitute directed credit for credit from the market. Second, while investment, and therefore total production, may go up even if the firm is not credit constrained, it will only go up if the firm has already fully substituted market credit with directed credit. They showed that bank lending and firm revenues went up for the newly targeted firms in the years when the priority sector was expanded to include them, and declined in the years where they were excluded again. They find no evidence that this was accompanied by substitution of bank credit for borrowing from the market and no evidence that revenue growth was confined to firms that had fully substituted bank credit for market borrowing. As already argued, the last two observations are inconsistent with the firms being unconstrained in their market borrowing.

The logic of credit constraints applies as much or more to human capital investments. Hart and

[22]The literature on the effects of cash-flow on investment is enormous. Fazzari, Hubbard and Petersen (1988) provide a useful introduction to this literature. The effects of family wealth on investment have also been extensively studied (see Blanchflower and Oswald (1998) for an interesting example). There is also a growing literature on the effects of community ties on investment (see, for example, Banerjee and Munshi (2004)).

Moore (1994), among others, have used human capital as the archetype of investment that cannot be collateralized, and therefore is hard to borrow against. This is made even more difficult by the fact that children would need to borrow for their education, or parents would need to borrow on their behalf. We return to this evidence below. The high responsiveness to user fees that we reviewed in section 2, and the evidence that investment in education are sensitive to parental income,[23] are both consistent with credit constraints. However, because human capital investments may involve direct utility or disutility (for example, a parent may like to see his child being educated), it is more difficult to come up with evidence that systematically nails the role of credit constraints for human capital investment. Edmonds (2004) is an interesting attempt to try to isolate the effect of credit constraints using household's response to an *anticipated* income shock. He studies the effect on child labor and education of a large old age pension program, introduced in South Africa at the end of the Apartheid. Many children live with older family members (often their grandparents). Women become eligible at age 60 and men become eligible at age 65. Since at the time he studies the program, the program was well in place and therefore fully anticipated, he argues that if more children start attending school as soon as their grandfather or grandmother crosses the age threshold and becomes eligible (rather than continuously, as they come closer to eligibility), this must be an indication of credit constraint. Indeed, he finds that child labor declines, and school enrollment increases, discretely when a household member becomes eligible.

• Summary

Credit constraints seem to be pervasive in developing countries. Of course, we are interested in whether the fact that access to capital varies across people helps us understand the productivity gap. If people invest different amounts because of differential access to capital, our intuitive presumption would be that capital is being misallocated, because there is no reason why richer people are always better at making use of the capital. This misallocation could be a source of difference in productivity. We will return to this question in section 5.

4.3 Problems in the Insurance Markets

Even if credit markets function well, and there is no limited liability, individuals may be reluctant to invest in any risky activity, for fear of losing their investment, if they are not properly insured against fluctuations in their incomes. Risk aversion leads to inefficient investment, and efficiency would improve with insurance (this idea is explored theoretically in Stiglitz (1969), Kanbur (1979), Kihlstrom and Laffont (1979), Banerjee and Newman (1991), Newman (1995) and Banerjee (2005)).

• Insurance in developing countries.

A considerable literature has investigated the extent of insurance in rural areas in developing countries

[23] See Strauss and Thomas (1995) for several studies along these lines.

(see Bardhan and Udry (1999) for a survey). Townsend (1994) used the ICRISAT data, a very detailed panel data set covering agricultural households in four villages in rural India to test for perfect insurance. The main idea behind this test is that with perfect insurance at the village level only aggregate (village-level) income fluctuation, and not idiosyncratic income fluctuations, should translate into fluctuation in individual consumption. He was unable to reject the hypothesis that the villagers insure each other to a considerable extent: Individual consumption seems to appear to be much less volatile than individual income, and to be uncorrelated with variations in income. This exercise had limits, however (see Ravallion and Chaudhuri (1997) for a comment on the original paper), and subsequent analyses, notably by Townsend himself, have shown the picture to be considerably more nuanced. Deaton (1997) shows that there is no evidence of insurance in Cote d'Ivoire. Townsend (1995) finds the same results across different areas in Thailand. Fafchamps and Lund (2003) find that, in the Philippines, households are much better insured against some shocks than against others. In particular, they seem to be poorly insured against health risk, a finding corroborated by Gertler and Gruber (2002) in Indonesia. Most interestingly, Townsend (1995) describes in detail how insurance arrangements differ across villages. While in one village there is a web of well-functioning risk-sharing institutions, the situations in other villages are different: In one village, the institutions exist but are dysfunctional; in another village, they are non-existent; finally, in a third village, close to the roads, there seems to be no risk-sharing whatsoever, even within family.

This last fact is attributed to the proximity to the city, which makes the village a less close-knit community, where enforcement of informal insurance contracts is more difficult. Coate and Ravallion (1993) was the first paper to build a theoretical model of insurance with limited commitment, and to show that, when the only incentive to contribute to the insurance scheme in good times is the fear of being cut away from the insurance in future periods, insurance will be limited. It will also be optimal to make payment contingent on past history, which will lead to a blur between credit and insurance (Ray (1998)). Udry (1990) presents evidence from Nigeria that is consistent with this model. The villages he studies are characterized by a dense network of loan exchange: Over the course of one year, 75% of the households had made loans, 65% had borrowed money, and 50% had been both borrowers and lenders. Ninety-seven percent of these loans took place between neighbors and relatives. Most importantly, the loans are "state-contingent": Both the repayment schedule and the amount repaid are affected by the lender's state *and* the borrower's state. This is evidence that credit is to some extent used as an insurance device. The resulting system is a mix of credit and insurance close to what the model of limited commitment would predict. However, and still consistent with this prediction, there is not enough of this "security" to fully insure households against income fluctuations: A shock to a particular borrower has a negative impact on the sum of the transfers received by his lender, which means that the lender did not fully diversify

risk.

Despite this evidence, we do not fully understand the reasons for the lack of insurance among households. It is unlikely that either limited commitment or the more traditional explanations in terms of moral hazard or adverse selection can explain why the level of insurance seems to vary from one village to the next, or why there is no more insurance against rainfall, for example.

• Consequences for investment.

Irrespective of the ultimate reason for the lack of insurance, it may lead households to use productive assets as buffer stocks and consumption smoothing devices, which would be a cause for inefficient investment. Rosenzweig and Wolpin (1993) argue that bullocks (which are an essential productive asset in agriculture) serve this purpose in rural India. Using the ICRISAT data, covering three villages in semi-arid areas in India, they show that bullocks, which constitute a large part of the households' liquid wealth (50% for the poorest farmers), are bought and sold quite frequently (86% of households had either bought or sold a bullock in the previous year, and a third of the household-year observations are characterized by a purchase or sale), and that sales tend to take place when profit realizations are low, while purchases take place when profit realizations are high. Since there is very little transaction in land, this suggests that bullocks are used for consumption smoothing. Because everybody needs bullocks around the same time, and bullocks are hard to rent out, Rosenzweig and Wolpin estimate that, in order to maximize production efficiency, each household should own exactly two bullocks at any given point in time. The data suggest that, for poor or mid-size farmers there is considerable underinvestment in bullocks, presumably because of the borrowing constraints and the inability to borrow and accumulate financial assets to smooth consumption: Almost half the households in any given year hold no bullock (most of the others own exactly two).[24] Using the estimates derived from a structural model where household use bullocks as a consumption smoothing device in an environment where bullocks cannot be rented and there is no financial asset available to smooth consumption, they simulate a policy in which the farmers are given a certain non-farm income of 500 rupees (which represents 20% of the mean household food consumption) every period. This policy would raise the average bullock holding to 1.56, and considerably reduce its variability, due to two effects: The income is less variable, and by increasing the income, it makes "prudent" farmers (farmers with declining absolute risk aversion) more willing to bear the agricultural risk.

Moreover, we observe only insurance against the risks that people have chosen to bear; the inability to smooth consumption against variation in income may lead households to choose technologies that are

[24]The fact that there is under-investment on average, and not only a set of people with too many bullocks and a set of people with too few, is probably due to the fact that bullocks are a lumpy investment, and owning more than two is very inefficient for production–there is no small adjustment possible at the margin.

less efficient, but also less risky. Banerjee and Newman (1991) argue, for example, that the availability of insurance in one location (the village), while its unavailability in another (the city), may lead to inefficient migration decisions, since some individuals with high potential in the city may prefer to stay in the village to remain insured.

There is empirical evidence that households' investment is affected by the lack of *ex post* insurance. Rosenzweig and Binswanger (1993) estimate profit functions for the ICRISAT villages, and look at how input choices are affected by variability in rainfall. They show that more variable rainfall affects input choices, and in particular, poor farmers make less efficient input choices in a risky environment. Specifically, a one standard deviation increase in the coefficient of variation of rainfall leads to a 35% reduction in the profit of poor farmers, 15% reduction in the profit of median farmers, and no reduction in the profit of rich farmers. Morduch (1993) specifically investigates how the anticipation of credit constraint affects the decision to invest in HYV seeds. Using a methodology inspired by Zeldes (1989), he splits the sample into two groups, one group of landholders who are expected to have the ability to smooth their consumption, and one group that owns little land, whom we expect *a priori* to be constrained. He finds that the more constrained group uses significantly less HYV seeds.

It is worth noting that the estimated impact of lack of insurance on investment is likely to be a serious underestimate. It is not clear how one could evaluate how much the lack of insurance affects investment. While we might observe certain options considered by the investor, there is no obvious way for knowing what other, even more lucrative choices, he chose not to even think about.

Another strategy for looking at the effects of underinsurance is to calculate the effect based on the assumption of specific utility function. This, in effect, is what Krussel and Smith (1998) do. They argue that for reasonable parameter values the effect on aggregate investment tends to quite small: This is because most people can self-insure quite well against idiosyncratic shocks, and those who cannot, mainly the very poor, do very little of the investment in any case. However as pointed out by a more recent paper by Angeletos and Calvet (2003), the Krusell and Smith result relies heavily on the assumption that one cannot limit exposure to risk by investing less. If investing more exposes you to more risk, even the non-poor will worry about risk, because they are the ones who invest a lot and therefore are exposed to a lot risk.

4.4 Local externalities

As we discussed in section 4, there is a line of work that focuses on coordination failures at the level of the economy: However, Durlauf (1993) shows that externalities do not have to be aggregated for the economy to exhibit multiple equilibria: Local complementarities (where adoption of a particular technology lowers production costs in a few "neighboring" sectors) can build up over time to affect aggregate behavior and

generate lower aggregate growth.

An example of strategic complementarity of this kind arises when agents are learning from each other. Banerjee (1992) shows how, when people try to infer the truth from other people's actions, this leads them to under-utilize their own information, and leads to "herd behavior". While this behavior is rational from the point of view of the individual, the resulting equilibrium is inefficient, and can lead to underinvestment, overinvestment, or investment in the wrong technology whatsoever.[25]

The impact of learning on technology adoption in agriculture has been studied particularly extensively. Besley and Case (1994) show that in India, adoption of HYV seeds by an individual is correlated with adoption among their neighbors. While this could be due to social learning, it could also be the case that common unobservable variables affect adoption of both the neighbors.[26] To partially address this problem, Foster and Rosenzweig (1995) focus on profitability. As we mentioned previously, during the early years of the green revolution, returns to HYV were uncertain and dependent on adequate use of fertilizer. In this context, the paper shows that profitability of HYV seeds increased with past experimentation, by either the farmers or others in the village. Farmers do not fully take this externality into account, and there is therefore underinvestment. In this environment, the diffusion of a new technology will be slow if one neighbors' outcomes are not informative about an individual's own conditions.[27] Indeed, Munshi (2004) shows that in India, HYV rice, which is characterized by much more varied conditions, displayed much less social learning than HYV wheat.

All of these results could still be biased in the presence of spatially correlated profitability shocks. Using detailed information about social interactions Conley and Udry (2003) distinguish geographical neighbors from "information neighbors", the set of individuals from whom an individual neighbor may learn about agriculture. They show that pineapple farmers in Ghana imitate the choices (of fertilizer quantity) of their information neighbors when these neighbors have a good shock, and move further away from these decisions when they have a bad shock. Conley and Udry try to rule out that this pattern is due to correlated shocks by observing that the choices made on an established crop (maize-cassava intercropping), for which there should be no learning, do not exhibit the same pattern.

The ideal experiment to identify social learning is to exogenously affect the choice of technology of a group of farmers and to follow subsequent adoption by themselves and their neighbors, or agricultural contacts. Duflo et al. (2003) performed such an experiment in Western Kenya, where less than 15% of the farmers use fertilizer on their maize crop (the main staple) in any given year despite the official recommendation (based on results from trials in experimental farms), as well as the high returns (in excess

[25] For a related model, see Bikhchandani, Hirshleifer and Welch (1992).

[26] See Manski (1993) for a discussion of the identification problem in social learning problems.

[27] Ellison and Fudenberg (1993) describe "rule of thumb" learning rules where individuals learn from others only if they are similar.

of 100%) that they estimated. They randomly selected a group of farmers and provided fertilizer and hybrid seeds sufficient for small demonstration plots in these farmers' fields. Field officers from an NGO working in the area guided the farmers throughout the trial, which was concluded by a debriefing session. In the next season, the adoption of fertilizer by these farmers increased by 17%, compared to the adoption of the comparison group. However, there is no evidence of any diffusion: People named by the treatment farmers as people they talk to about agriculture did not adopt fertilizer any more than the contacts of the comparison group. The neighbors of the treatment group actually tended to adopt fertilizer *less* often, relative to the neighbors of the comparison group. This is not because only experimentation in one's own field changes someone's priors: When randomly selected friends were invited to attend the harvest, the debriefing session, and other key periods of the trials, they were as likely to adopt fertilizer as the farmers who participated in the experiment. Rather, it suggests that, spontaneously, information about agriculture is not shared. This points towards another type of externality and source of multiple steady states: When there is very little innovation in a sector, there is no news to exchange, and people do not discuss agriculture. As a result, innovation dies out before spreading, and no innovation survives.

Depending on the priors of the individuals, social learning can either decrease or increase investment. In Kenya, Miguel and Kremer (2003) show that random variation in the number of friends of a child who was given the deworming medicine had a *negative* impact of the propensity of a child to take the medicine. They attribute this to the fact that parents may have initially over-estimated the benefits of the deworming drug.

In addition to social learning, there are many other sources of local interactions. First, people imitate each other even when they are not trying to learn, because of fashion or social pressure. Social norms may prevent the adoption of new technologies, because coordinating on a new equilibrium may require many people to change their practices at the same time.[28] Second, there are several sources of positive spillovers between industries located close to each other. Silicon Valley-style geographic agglomerations occur in the developing world as well, such as the software industry in Bangalore. Ellison and Glaeser (1997) show that, in the U.S., most industries are indeed more concentrated than they would be if firms decided to place their plants randomly. Only about half of this concentration is explained by the fact that some locations have natural advantages for specific industries (Ellison and Glaeser (1999)).

In addition to the traditional arguments for positive spillovers, such as transport costs (fast telecommunication lines that were installed for the software industry in Bangalore greatly reduced the cost of setting up call centers, for example), intellectual spillovers or labor market pooling, a powerful reason for geographical agglomeration in developing countries is the role of a town's reputation in the world market. For example, outsiders who want to start working in garment manufacturing come to Tirupur,

[28]See Munshi and Myaux (2002) for an example on the spread of family planning in Bangladesh.

the small town studied in Banerjee and Munshi (2004), despite their difficulty of finding credit there, because this is the place where large American stores come to place orders. There is a sense in which the town has a good reputation, for quality and timeliness of the delivery, and everybody who works there benefits from it. Tirole (1996) models "collective reputation": If many people in a group are known to deliver good quality products, buyers will have high expectations and be willing to trust the sellers to produce more elaborate products, where quality matters. In turn, this will encourage sellers to produce high quality products to avoid being outcast from the group, which will sustain a "high quality-high trust equilibrium". But if buyers are expected to only ask for basic products in the future, building a reputation for high quality is not useful, and opportunistic sellers will produce low quality in the first period. Knowing this, sellers indeed have the incentive to ask for simple products, and the bad equilibrium persists. In this world, history matters. A collective reputation for low quality is very difficult to reverse, and a collective reputation for high quality is valuable. We should therefore expect groups to try to set up institutions to develop a good collective reputation. There is certainly some indication that this is happening. For example, the association of Indian software firms (NASSCOM) tries to help the firms access quality certifications such as ISO 9001, SEI, or others. Much more work on whether collective reputation matters in practice is, however, clearly needed before we can assess the empirical relevance of these sources of externality.

To summarize, externalities can explain very large variations in productivity and investment rates across otherwise similar environments.

4.5 The Family: Incomplete Contracts Within and Across Generations

Investment in human capital often pays in the long term, and in many crucial instances must be done by parents on behalf of the child. In this context, the way the decisions are made in the family has a direct impact on investment decisions. In the benchmark neo-classical model (Barro (1974); Becker (1981)), parents value the utility of their children, perhaps at some discounted rate. This world tends to be observationally equivalent to one where an individual maximizes his long run income, and has the same strong convergence properties. However, if parents are not perfectly altruistic, the ability to constrain the repayment of future generations influences investment decisions. Banerjee (2004) studies the short and long run implications of different ways to model the family decision-making process. He shows that incomplete contracting between generations generates potentially large deviations from the very strong convergence property of the Barro-Becker model. Deviations also occur if parents value human capital investment for its own sake (for example, because people like to see their children happy).[29]

[29] Part of the reason why investment in human capital may appear like a preference factor is that individuals want their offspring to thrive and survive. In the U.S., Case and Paxson (2001) and Case, Anne and McLanahan (2000) find that

In particular, even with perfect credit markets, parental wealth will determine how much is invested in human capital. There can be more than one steady state, and there can be inequality in equilibrium. In this world, increases in returns to human capital may not lead to an increase in human capital, if the production of human capital is skill-intensive (the increase in the price of teachers may dominate the added incentives to invest in education).

Many studies have shown that human capital investment is correlated with family income (see Strauss and Thomas (1995) for references for developing countries). In general, however, it is difficult to separate out the pure income effect from the effect of an increase in the returns to investing in human capital, differences in the opportunity cost or the direct cost of schooling, and different discount rates. For example, in the Barro-Becker model, families with a lower discount rate will tend to be richer and more likely to invest in education. To avoid this problem, a few studies have focused on exogenous changes in government transfers. For example, Carvalho (2000) shows that an increase in pension income in Brazil led to a decrease in child labor and an increase in school enrollment. Duflo (2003) shows that, in South Africa, girls (though not boys) have better nutritional status (they are taller and heavier) in households where a grandmother is the recipient of a generous old age pension program.

This paper also touches on another set of issues. Different members of the family may have different preferences. If education and health were pure investment, and if the members of the household bargained efficiently (as in Lundberg and Pollack (1994, 1996) or the papers reviewed in Bourguignon and Chiappori (1992)), this would not have any impact on education or health decisions. However, if either assumption is violated, it means that not only the size of the income effects, but who gets the income, will affect investment decisions. In the case of the South African pensions, this was clearly the case: Pensions received by men had no impact on the nutritional status of children of either gender. This may come from the fact that women and men value child health differently, or from the fact that the household is not efficient, and a specific individual is more likely to invest in children if the returns are more likely to directly accrue to her.

If the household does not bargain efficiently, the consequences extend beyond investment in human capital to all investment decisions. In a Pareto efficient household, production and consumption decisions are separable: The household should choose inputs and investment levels to maximize production, and then bargain over the division of the surplus. This property will be violated if individuals make investment decisions with an eye toward maximizing the share of income that directly accrues to them. Udry (1996) shows that, in Burkina Faso, after controlling for various measures of the productivity of the field (soil

investment in children is lower when they do not live with their birth mother. Using data from several African countries, Case, Paxson and Ableidinger (2002) find that the gap between the probably of being enrolled in school for orphans and non-orphans can be in part accounted for by the fact that they are less likely to live with at least one parent, and more likely to live with non-relatives.

quality, exposure, slope, etc.), crop, year, and household fixed effects, yields on plots controlled by women are 20% smaller than yields on men's plots.[30] This does not seem to be due to the fact that women and men have different production functions. Instead, this difference is largely attributed to differences in input intensity: In particular, much less male labor and fertilizer is used on plots controlled by women than on plots controlled by males. The fertilizer result is particularly striking, since there is ample evidence that it has sharply decreasing returns to scale. Udry estimates that the households could increase production by 6% just by reallocating factors of production within the household.

Udry explains underinvestment on women's plots by their fear of being expropriated by their husband if he provides too much labor and inputs. Another reason for inefficient investment may be the fear of being fully taxed by family members once the investment bears fruit. Again, an efficient household would first maximize production. However, the specific claims that a household member (or a neighbor, or a member of the extended family) can make on someone's income stream may lead to inefficient investment. Consider, for example, a situation where individuals have the right to make emergency claims on the income or savings of others in their group (for example, if someone is sick and has no money to pay for the doctors, others in his extended family have an obligation to pay the doctor). Consider a savings opportunity that will increase income by a large amount in the future (for example, saving money after harvest to be able to buy fertilizer at the time of planting). If everybody could commit not to exercise their claim during the period where the income needs to be saved, the money should be saved, and the proceeds eventually distributed to those who have a claim on it, and everybody would be better off. However, if no such commitment is possible, the individual who earned the income knows that it is likely that, should he choose to save enough for fertilizer, a claim will be exercised in the period during which the money needs to be held. He is then better off spending the money right away: Even if individuals are rational and have a low discount rate, as a group they will behave as "hyperbolic discounters", who discount the immediate future relative to today more than future periods relative to each other (Laibson (1991)). The level of investment will be low in the absence of savings opportunities offering some commitment to household members.

The fact that investments are often decided within a family, rather than by a single individual, or that the proceeds of the investment will be shared among a set of people who have not necessarily supported the cost of the investment therefore greatly complicates the incentive to invest. This may, once again, explain why some potential investments with high marginal product are not taken advantage of. It is worth noting that the lack of credit and insurance in poor countries makes these problems particularly acute there. For example, the lack of credit markets means that investment decisions are taken within the

[30] In Burkina Faso, as in many other African countries, agricultural production is carried out simultaneously on different plots controlled by different members of the household.

families–e.g., women cannot borrow to get the optimal amount of fertilizer on their plot–and the lack of insurance plays an important role in justifying the norms on family solidarity that seem to be hindering productive investment.

4.6 Behavioral Issues

Individuals in the developing world appear not only to be credit constrained, but also to be savings constrained. Aportela (1998) shows that when the Mexican Savings Institution "Pahnal" (Patronoto del Ahorro Nacional) expanded its number of branches through post offices in poor areas and introduced new savings instruments in the 1990s, household's savings rates increased by 3% to 5% in areas where the expansion took place. The largest increase occurred for low income households.

When an individual (or his household) has time-inconsistent preferences, formal savings instruments may increase savings rates even when they offer very low returns (even compared to holding onto cash), because they offer a commitment mechanism. Micro-credit programs may also be understood as programs helping individuals to commit to regular reimbursements. This is particularly clear for programs, like the FINCA program in Latin America, which require that their clients maintain a positive savings balance even when they borrow.[31]

Duflo et al. (2003) provide direct evidence that there is an unmet demand for commitment savings opportunities among Kenyan farmers, and that investment in fertilizer increases when households have access to this opportunity. In several successive seasons, they offered farmers the option to purchase a voucher for fertilizer right after harvest (when farmers are relatively well off). The vouchers could be redeemed for fertilizer at the time when it is necessary to plant it. The take up of this program was quite high: 15% of the farmers took up the program the first time it was tried with farmers who had never encountered the NGO before. Net adoption of fertilizer increased in this group. The program was then offered to some of the farmers who had participated in the pilot program mentioned above (and thus had the opportunity to test the fertilizer for themselves, and trusted the NGO), and in this group, the take up was 80%. There is also direct evidence of the difficulty for farmers to hold on to cash: In other experiments, when farmers were given a few days before they could purchase the voucher, the take up fell by more than 50%. When they were offered the option of having the fertilizer delivered at home at the time they actually needed it (and to pay for it then), none of the farmers who had initially signed up for the program had the money to pay for the fertilizer when it was delivered.

This area of research is quite recent, and wide open. Many questions need answering, and the

[31] Karlan (2003) argues that simultaneous borrowing and savings by many clients in these institutions can be explained by the value to the small business owner of the fixed repayment schedule as a discipline device. Gugerty (2000) and Anderson and Baland (2002) interpret rotating credit and savings (ROSCAs) institutions in this light.

area of applicability is wide. For example, what is the best way to increase parents' willingness to invest in deworming drugs? Why don't all parents sign the authorization form which will grant free access to deworming to their children (Miguel and Kremer (2003))? Is it a rational decision or is it procrastination? Why does the take up of the deworming drug fall so rapidly when a small cost-sharing fee is introduced (Miguel and Kremer (2003))? Understanding the psychological factors that constrain investment decisions, and the role that social norms play in disciplining individuals, but also potentially in limiting their options, is an important area for future research. Several randomized evaluations are trying to make progress in this area. They are addressing questions as diverse as: What is the role of marketing factors in the access of poor people to loans in South Africa (Bertrand, Karlan, Mullainathan and Zinman (2004))? Do poor people take advantage of savings products with commitment options in the Philippines (Ashraf, Karlan and Yin (2004))? What prevents people from doing a small action that would lead them to a high return (Duflo et al. (2003))? What factors (deadline, framing, etc.) make it more likely they will take an action (Bertrand et al. (2004))?

A defining characteristic of these projects is that they do not involve laboratory experiments: Like the research on fertilizer in Kenya, they set up real programmes which are likely to increase poor people's investment and improve welfare if they indeed deviate from perfect rationality in the way the psychological literature suggests. In order to be fruitful, this agenda will need to avoid simply transplanting to developing countries some of the insights developed by observing behaviors in rich countries. Being poor almost certainly affects the *way* people think and decide. Decisions, after all, are based not on actual returns but on what people perceive the returns to be, and these perceptions may very well be colored by their life experience. Also, when choices involve the subsistence of one's family, trade-offs are distorted in different ways than when the question is how much money one will enjoy at retirement. Pressure by extended family members or neighbors is also stronger when they are at risk of starvation. It is also plausible that decision-making is influenced by stress. What is needed is a theory of how poverty influences decision making, not only by affecting the constraints, but by changing the decision making process itself.[32] That theory can then guide a new round of empirical research, both observational and experimental.

5 Calibrating the Impact of the Misallocation of Capital

In this long list of potentially distorting factors there are some, like government failures or credit market failures, that most people find *a priori* plausible, and others, such as intra-family inefficiencies or learning externalities that are more contentious, and yet others, like the behavioral factors, that have not yet been

[32] See Ray (2003) for a very nice attempt to start in this direction.

widely studied. However, even where the prima facie evidence is the strongest, we cannot automatically conclude that the particular distortion has resulted in a *significant* loss in productivity.

To get a sense of the potential productivity loss, we return to the Indo-U.S. comparison. Taking as given the stock of capital in India and the U.S. today, *any of the multiple distortions listed above* could have affected productivity in two different ways: First, there may be across-the-board inefficiency, because everyone could have chosen the wrong technology or the wrong product mix. Second, capital may be misallocated across firms: There may be differences in productivity across firms, either because of differences in scale, or because of differences in technology or because some entrepreneurs are more skilled than others, and the distribution of capital across these firms may be sub-optimal, in the sense that the most productive firms are too small.

Here we have chosen to emphasize this latter source of inefficiency, motivated in part by the evidence, discussed above, that tells us that there is enormous differences in productivity across firms. *We take no stance on how such an inefficient allocation of capital came about, nor on why the firms do not make the right choices, either of scale or or technologies.* Lack of access to credit is, of course, a potential explanation for both, but it could equally be explained by lack of insurance, the fear of confiscation by the government, or the gap between real and perceived returns.

The goal of this section is to set-up and calibrate a simple model, to investigate whether the misallocation of capital across firms within a country *can* explain the aggregate puzzles we started from: the low output per worker in developing countries, given the level of capital, and the low marginal product of capital, given the output per worker. We do not claim that we necessarily have the right explanation; there are simply too many degrees of freedom in this kind of exercise. Nevertheless we feel that the exercise has some value, not least because it gives us some reason to believe that we have not been entirely misguided in emphasizing the role of misallocation. Moreover, as we will see, it does rule some kinds of misallocation stories in favor of others.

We begin with a model where the misallocation only affects the scale of production, because all the firms share the same technology. Scale obviously does not matter where there are constant returns to scale, so we need to turn to a model where there are diminishing returns at the firm level.[33] We will show that, with realistic assumptions about the relative firm size in India and the U.S., this model cannot go very far in explaining the aggregate facts. We then turn to a model where a better technology can be purchased for a fixed cost. We show that this model, coupled with the misallocation of capital, will help generate the aggregate facts, with realistic assumptions about the distribution of firm sizes.

[33]The obvious alternative –increasing returns at the firm level– will clearly not fit the basic fact that there is more than one firm in the U.S., or that the marginal product of capital is higher in India than in the U.S.

5.1 A Model with Diminishing Returns

• Model set-up

Consider a model where there is a single technology that exhibits diminishing returns at the firm level, say, $Y = AL^{\gamma}K^{\alpha}$, with $\gamma < 1 - \alpha$. Also, we will assume that the economy has a fixed number of firms: Without that assumption, everyone will set up multiple minuscule firms, thereby eliminating the diminishing returns effect. To justify this we make the standard assumption that the economy has a fixed number of entrepreneurs and each firm needs an entrepreneur.

Under these assumptions every firm would invest the same amount when markets function perfectly, but when different firms are of different sizes, the marginal product would vary across the firms and efficiency would suffer. The question is whether these effects are large enough to help us explain what we see in the data. Given that we are still within the class of Cobb-Douglas models, we know from section 3 that we cannot both get the right ratio for output per worker and the right value of the ratio of marginal products; we therefore focus on the output per worker.

We start with a population of firms indexed by i, and that firms face a limit on how much they can borrow, so that for firm i, $K \leq K(i)$. The demand for labor from a firm that invests $K(i)$, is given by $\left[\frac{A\gamma K(i)^{\alpha}}{w}\right]^{\frac{1}{1-\gamma}}$. We assume a perfect labor market, so that given the level of capital, labor is efficiently allocated across firms. Labor market clearing then requires that

$$w = A\gamma \left[\frac{\int [K(i)^{\alpha}]^{\frac{1}{1-\gamma}}\, dG(i)}{\overline{L}}\right]^{1-\gamma},$$

where $G(i)$ represents the distribution of i and $\overline{L}$ is labor supply per firm. Since wages are a fraction γ of output-per-worker, it follows that output-per-worker will be

$$A\left[\frac{\int [K(i)^{\alpha}]^{\frac{1}{1-\gamma}}\, dG(i)}{\overline{L}}\right]^{1-\gamma}.$$

Consider an economy where, for any of the reasons we have outlined above, some firms have access to more capital than others. In particular, assume that in equilibrium a fraction λ of firms get to invest an amount K^1 and the rest get to invest $K^2 > K^1$.[34] This would clearly explain why the marginal product of capital varies within the same economy. We would also expect that this inefficiency in the allocation of capital would lower productivity relative to the case where capital was optimally allocated. To get at the magnitude of the efficiency loss, note that output per worker in this economy will be:

$$A[\frac{\lambda(K^1)^{\alpha/(1-\gamma)} + (1-\lambda)(K^2)^{\alpha/(1-\gamma)}}{\overline{L}}]^{1-\gamma}.$$

[34] Since all firms face the same technology and there are diminishing returns to scale, this would not happen in the absence of these imperfections (all the firms should invest the same amount, $\lambda K^1 + (1-\lambda)K^2 = \overline{K}$).

We compare this economy with another which has a TFP of A', a labor force $\overline{L}'$ and a capital stock $\overline{K}'$, which is, in contrast with the other economy, allocated optimally across firms. To say something about productivity we also need to say how many firms there are in this economy. Let us start by assuming that the number of firms is the same. Then the ratio of output-per-worker in our first economy to that in the second is:

$$(A/A')(\frac{\overline{K}}{\overline{L}}/\frac{\overline{K}'}{\overline{L}'})^{\alpha}(\overline{L}'/\overline{L})^{1-\alpha-\gamma}[(\lambda(K^1/\overline{K})^{\alpha/(1-\gamma)}+(1-\lambda)(K^2/\overline{K})^{\alpha/(1-\gamma)}]^{1-\gamma}. \tag{6}$$

We already noted that for the India-U.S. comparison, the ratio $\frac{\overline{K}}{\overline{L}}/\frac{\overline{K}'}{\overline{L}'}$ is about 1: 18. The same source (the Penn World Tables) tells us that $\overline{L}/\overline{L}'$ is about 2.7. What are reasonable values of α and γ? For $1-\alpha-\gamma$, which is the share of pure profits in the economy, we assume 20%, which is what Jovanovic and Rousseau (2003) find for the U.S. This is presumably counted as capital income, so we keep $\gamma = 0.6$ and set $\alpha = 0.2$.

First consider the case where $\lambda = 1$, so that capital is efficiently allocated in both countries. Then the productivity ratio ought to be $(A/A')(\frac{\overline{K}}{\overline{L}}/\frac{\overline{K}'}{\overline{L}'})^{\alpha}(\overline{L}'/\overline{L})^{1-\alpha-\gamma}$: Assuming that $2A = A'$, as before, because of the human capital differences across these economies, the ratio works out to be $\frac{1}{2}\left(\frac{1}{18}\right)^{0.2}(\frac{1}{2.7})^{0.2} = 23\%$. Recall from equation 5 that the model with constant returns predicted that output should be 6.35 higher in the U.S., or, equivalently that output per capita in India should be 15.7% the U.S. level. The 23% predicted by the current model is, of course, even further from the 9% we find in the data. The reason why this model does worse is because the production function is more concave: The concavity penalizes the U.S., which has more capital relative to India.

- What if capital is misallocated?

To bring in the effects of misallocating capital, we need to determine the size of the gap between K^2 and K^1 that we can reasonably assume. One way to calibrate these numbers is to make use of the estimate of Banerjee and Duflo (2004) that in India, there are firms where the marginal product of capital seems to be close to 100%. On the other hand, some seem to have access to capital at 9% or so, and therefore may well have a marginal product reasonably close to 9% (Timberg and Aiyar (1984)). From the production function, we know that if we assume that K^1 corresponds to the firm with a marginal product of 100%, while K^2 is the firm with the marginal product of 9%, then $(K^2/K^1)^{\frac{\alpha}{1-\gamma}-1} = \frac{9}{100}$ or $K^2/K^1 = (\frac{100}{9})^2 = 123$. We can now evaluate the ratio of output-per-worker in the two economies for any given value of λ, the fraction for firms with capital stock K^1. To pin down λ, we use the fact that the average of the marginal product in India seems to be somewhere in the range of 22%. In our model, under the assumption that the marginal dollar is allocated between small firms and large firms in the

same proportion as the average dollar, the average marginal product of capital is given by:

$$\frac{\lambda}{\lambda + 123(1-\lambda)}100 + \frac{(1-\lambda)123}{\lambda + 123(1-\lambda)}9.$$

Since this is equal to 22% we have that $\lambda = 0.95$. We can now compute the extent of productivity loss due to the misallocation. From equation 6, this is given by the expression $[(\lambda(K^1/\overline{K})^{\alpha/(1-\gamma)} + (1-\lambda)(K^2/\overline{K})^{\alpha/(1-\gamma)}]^{1-\gamma}$ in equation 6. Under the assumed values, it is approximately 0.8. In other words, the misallocation brings the productivity ratio we expect to see between India and the U.S. down from 23% to about 18%.

Relative to the neo-classical model we started from (which generates an output per worker in India of 15.7% of the U.S. level), moving to this model therefore does not help close the productivity gap between India and the U.S. The problem is, once again, that the additional productivity gap that the misallocation generates is more than compensated for by the effect of making the production function concave while keeping the number of firms fixed.

- Can we do better than this?

We could try to make the misallocation problem worse in India by making the big firms bigger and the small firms smaller. However the industry structure we started with was already rather extreme. Table 5.1 lists, for nine of the largest industries in India (where industry is defined at 3 digit level) outside of agriculture, known for having a substantial presence of small enterprises, some measures of variation in firm sizes (where size is measured by the net fixed capital in year 2000) from the Annual Survey of Industries (ASI). In the industry described by our model, the large firm in our model is 123 times the size of the small firm whereas in the ASI data, even the 95th percentile firm in the median industry is no more than 72 times the 25th percentile firm. The firm that is $\frac{1}{123}$ times the 95th percentile firm in the median industry is around the 20th percentile in the size distribution. More than 50% of the capital stock in the Indian economy is in firms that are bigger than the "small" firm and smaller than the "large" firms as we defined them here. Realism therefore requires that we move weight *away from the extremes,* but this will not help us fit the data, since it dampens the concavity effect that is at the heart of our argument.

Another problem with making the big firms bigger is that the big firms in our data are already too big, relative to their American counterparts: Since $K^2 = 123K^1$, $\overline{K_I} = [0.955 + 123(0.045)]K^1 = 6.5K^1$ and $K^2 \approx 19\overline{K_I}$. Now since $\left(\frac{\overline{K}}{\overline{L}}\right)_I / \left(\frac{\overline{K}}{\overline{L}}\right)_U$ is about 1/18 and $\overline{L}_I/\overline{L}_U$ is about 2.7, $\overline{K}_I/\overline{K}_U = 0.15$. Therefore $K^2/\overline{K}_U = 2.85$. The biggest US firms in our model are a third of the biggest Indian firms, whereas it is typically assumed that big US firms are, if anything, bigger than big Indian firms.

Finally note that in our model, the ratio of the 95th percentile firm to the 5th percentile firm in the

Table 1: Distribution of firm size (Annual Survey of Industry, 2000)

	95-5 ratio	median-5 ratio	mean-5 ratio	5th percentile
Manufacture of Pasteurized Milk	1007	95	216	61466
Flour milling	786	150	285	29899
Rice milling	1392	90	620	5681
Cotton Spinning	22300	440	5423	12870
Cotton Weaving	3093	31	1292	14159
Textile garment manufacture	1581	104	410	22461
Curing raw hides and skins	235	10	53	37075
Manufacture of footwear	2639	122	683	21825
Manufacture of car parts	1700	29	504	84103

median industry is approximately 1,600:1.[35] Given the production function, we know that the marginal return on capital in the two firms should differ by a factor of $1600^{1/2} = 40 : 1$. If the biggest firms pay about 9% for their capital, the smaller firms must have a marginal product that exceeds 360%, which seems implausible. Making the gap between small and large firms even larger would clearly exacerbate this problem.

Another possible way to augment the productivity gap is to give up the assumption that the two economies have same number of firms. Suppose the U.S. had $\lambda > 1$ times as many firms as India: Then the labor productivity ratio computed above would have to be divided by $\lambda^{1-\alpha-\gamma}$. If λ were equal to 32, the ratio of labor productivity in India to that in the U.S. would be 9%, which is what we find in the data.

Of course, increasing the number of the firms in the U.S. will tend to make the average firm in the U.S. smaller: Even with the same number of firms in the two countries, the fact that the biggest firms in India have about 18 times the average capital stock means that they are about 3 times the size of U.S. firms, which seems implausible. If there are 32 times as many firms in the U.S., the average U.S. firm would be about a 1/100th of the biggest Indian firm, close to 25% in the Indian size distribution. This seems entirely counterfactual.

- Taking Stock

To sum up, moving to this more sophisticated model does not help us fit the macro facts better. It obviously does suggest a simple theory of the cross-sectional variation in returns to capital, which

[35] The median industry is the Textile Garment Manufacturing industry.

is entirely absent from the model with constant returns, but ends up throwing up a number of other problems that a theory of this type will need to deal with. In particular, a successful explanation has to be consistent with the fact that the firm size distribution in India has a large part of its weight near the mean/median; that even the biggest Indian firms are not larger than the bigger US firms in the same industry; and that the marginal product of capital in the average small Indian firm, while large (even 100%) is probably not 300% or more.

The next section introduces an alternative model where firms differ both in scale and in technology, but still retains the assumption that there is no inherent difference between these alternative investors.

5.2 A Model with Fixed Costs

• Model Set up

Consider a world where setting up requires a fixed start-up cost in addition to an entrepreneur, but once these are in place, capital and labor get combined as in a standard Cobb-Douglas with diminishing returns. This fixed cost could come from many sources: Machines come in certain discrete sizes and even the smallest machine may be expensive from someone's point of view. Buildings, likewise, are somewhat indivisible, at least by the time we come down to a single room or less. Marketing and building a reputation may also requires an indivisible up-front investment—Banerjee and Duflo (2000) describe the costs that a new firm in the customized software industry has to pay in terms of harsh contractual terms, until it has a secured reputation. Turning to investment in human capital, it also appears that the first five years or so of education may have much lower returns than the next few years, which in effect makes the first few years of education a fixed cost.[36] Finally, as emphasized by Banerjee (2003) the fixed cost may be in the financial contracting that the firm has to go through—starting loans are often expensive because the lender cannot trust the borrower with a big loan and when the loan is small, the fixed costs of setting up the contract loom large.

Formally, we assume a production function $y = A(K - \underline{K})^{\alpha} L^{\gamma}$. Since we continue to assume that the firm can buy as much labor as it wants, the production function can be rewritten as:

$$A^{\frac{1}{1-\gamma}} \left[\frac{\gamma}{w}\right]^{\frac{\gamma}{1-\gamma}} [K - \underline{K}]^{\frac{\alpha}{1-\gamma}}. \tag{7}$$

We continue to assume that $\gamma + \alpha < 1$, so that there are diminishing returns. The average cost function in this world has the classic Marshallian shape: Average costs go down first as the fixed costs

[36] All the estimates (14) we could find of Mincerian returns at different levels of education suggest that, in developing countries, the marginal benefits of a year of education increases with the level of education (in the U.S., it appears to be very flat). Schultz (2004) finds the same result in his study of six African countries.

get amortized over more and more output and then start to rise again. The optimal scale of production is given by the equality of the marginal and average product of capital, which reduces to:

$$K = \underline{K}\frac{1-\gamma}{1-\gamma-\alpha}.$$

We allow firms the option of choosing between alternative technologies. Assume that there are three alternative technologies available, characterized by three different levels of the fixed cost, $\underline{K}_1$, $\underline{K}_2$ and $\underline{K}_3$, three differing levels of labor and capital intensity, $\{(\alpha_1, \gamma_1), (\alpha_2, \gamma_2), (\alpha_3, \gamma_3)\}$ and three correspondingly different levels of productivity, A_1, A_2 and A_3. We make the usual assumption that a higher cost buys a higher levels of TFP, i.e., that $\underline{K}_1 \leq \underline{K}_2 \leq \underline{K}_3$ and $A_1 \leq A_2 \leq A_3$.

Compared to a Cobb-Douglas model with diminishing returns, this formulation has a number of advantages. First, it allows firms to have large differences in size without necessarily large differences in the marginal product of capital, since they could be using different technologies. The fact that there are firms in the same industry operating at very different scales posed a problem for the model with diminishing returns because the implied variation in the marginal product of capital seems implausibly large. Second, the fact that a lot of the capital stock in India is in firms that are close to the mean, was problem when we had global diminishing returns, because with diminishing returns, firms close to the mean are at the optimal scale. In our present model, the right scale for Indian firms is actually much bigger than the mean. A part of the inefficiency comes precisely from the fact that there are many firms that are concentrated near the mean. Third, as noted above, this model generates a unique optimal scale of production, which would provide a reason why the biggest (and most productive) Indian and U.S. firms could be more or less the same size. Fourth, because efficient firms tend to be quite large, it is easy to see why India, with its multitude of firms that are too small, will be inefficient relative to the US, where all firms are at the right scale. Fifth, the fact that production requires a fixed cost helps explain why, despite the diminishing returns from technology, we do not see people setting up a very large number of very small firms, thereby completely eliminating the diminishing returns effect. In this case, we can let the number of firms be determined by what people are willing to invest, in combination with what we know about the fixed costs (actually as noted below, we cheat slightly on this point, but only because it simplifies the calculations). Sixth, the fact that we allow the number of firms to be determined endogenously means that there is less overall diminishing returns. When we compare the U.S. and India, this helps explain why the productivity gap is so large and why interest rates are not lower in the U.S. Finally making this assumption alters the nature of the link between the marginal product of capital and its average product. With a Cobb-Douglas, the ratio of the average product is always proportional to the marginal product. Here, the average product starts lower than the marginal product but grows faster and eventually becomes larger than the marginal product. In other words, as firm size goes up the

ratio of the marginal product of capital to its average product goes down, at least initially. This would suggest that the ratio of the average products of capital in India and the U.S., should be less than the ratio of the marginal products, and indeed we find that while output-per-worker is 11 times larger in the U.S., capital-per-worker is 18 times as large, implying an average product ratio of about 1.6:1, as against the 2.5:1 ratio of marginal products delivered by the standard Cobb-Douglas model. This is clearly an *a priori* advantage of this formulation, since, as we noted in section 3, the proportionality between the average product and the marginal product prevents any model based on a Cobb-Douglas production function to fit these facts.

Interestingly, this model brings together elements—market imperfections and some increasing returns—that are also being invoked by recent work in new growth theory (Aghion et al. (2004), for example) for the same purpose, namely to explain the lack of convergence. However, the increasing returns and the credit constraints here are at the level of the firm, whereas in the aggregative growth literature they are at the level of the economy. Indeed, if there is no misallocation and no lack of people to start new firms, the aggregate production function generated by our economy exhibits constant returns in labor and capital: Indeed it has exactly the form that Lucas started with: $Y = AK^{\alpha}L^{1-\alpha}$.

In order to impose restrictions on the parameters of the model, we make use of the industry data described in table 5.1. We describe the representative Indian industry by a 3-point distribution of firms sizes, with fractions λ_1, λ_2, and λ_3 at K_1, K_2 and K_3. The first group of firm is made of the bottom 10% of the the distribution of firms, and we assigned to them the size of the firm at the 5th percentile of the actual size distribution in the data. Likewise, we assume that the top 10% of all firms are in the group of "large firm", and that their size is that of the firm at the 95th percentile of the firm size distribution.[37] The rest we assign to the middle category, whose size we set at the mean for the distribution. We assume that the largest firm is 1,600 times as big as the smallest firm, which is roughly the median value of these ratios across these nine industries in our data.

These parameter values imply that the mean firm size in the industry will be 800 times as large as the 5th percentile firm, which is higher than the mean in the median industry in our data (500 times), but well within the existing range in the data. Once again we are interested in the within-economy variation in returns to capital. We therefore assume, as before, that the small firms have a marginal product of 100% while the medium sized firms have a marginal product of just 9%.

The more unorthodox assumption is that the large firms also have a marginal product of 100%. While clearly somewhat artificial, this is meant to capture the idea that the best technology is expensive and only the biggest firms in India can afford to be at the cutting edge, an idea that is very much in the spirit

[37] We pick the 5th and the 95th percentile to make the difference in the returns to capital between the biggest and smallest firms as large as possible.

of the McKinsey Global Institute's study of a number of specific Indian industries. However, they are still relatively small and therefore the marginal returns on an extra dollar of investment are very high. The rest of the firms use cheaper (i.e., lower $\underline{K}$) but less effective technologies. In particular, the small firms are simply too small (which explains their high marginal product), and the middle category consists of firms that have exhausted the potential of the mediocre technology that they can afford but are too small to make use of the ideal technology.

How plausible is our assumption about industry structure? The average capital stock of the 95th percentile firms in the median industry was Rs. 36 million, which puts them at a size just above the category of firms that are the focus of Banerjee and Duflo (2004). The point of that paper was that a subset of these firms (the firms that attracted the extra credit after the policy change) had marginal returns on capital of close to 100%. Therefore it is not absurd to assume that the large firms in our model economy have very high returns. Once we accept the idea that some large firms are very productive, given that the average marginal product is probably close to 22%, it is obviously very likely that there are many smaller firms that have a *lower* marginal product than the largest firms. Indeed, when we calculate the average marginal product based on our assumptions, under the premise that the marginal dollar is distributed across the three size categories in the ratio of their share in the capital stock, the average marginal product turns out to be about 27%.

Even with this long list of assumptions, we do not have enough information to compute output per worker in our model economy—there are several remaining degrees of freedom. First, we need to choose units: Our assumption, which simplifies calculations, is that capital is measured in multiples of the small firm. Finally, we assume that $\underline{K}_1 = 0$, $\underline{K}_2 = 100$, and $\underline{K}_3 = 800$. The assumption that $\underline{K}_3 = 800$, implies that the biggest Indian firms (which have 1,600 units of capital) are operating at the bottom of the average cost curve—given by $\underline{K}\frac{1-\gamma}{1-\gamma-\alpha}$.[38]

• Results: Output per worker and average marginal product of capital

Under these assumptions, we can use the assumed marginal products to solve for A_1, A_2 and A_3. According to these calculations, TFP in the medium firms is about 1.4 times bigger than that in the small firms, and that in the large firms is about 2.7 times that in the medium firms. Nevertheless, given the assumed limits on how much they can invest, each category of firms is optimizing by choosing its current technology. However, large gains in productivity are obviously possible if the economy can reallocate its capital so that all the firms adopt the most productive technology.

[38] This is where we cheat, since with decreasing returns to scale, there could again be an infinity of very small firms, so that all the firms should be in the small group. We can prevent that if we assume that the smallest feasible firm size is actually ϵ greater than zero, and only a certain number of entrepreneurs are able (or willing) to invest at least ϵ.

To see how large this gain may be, we do another India-U.S. comparison. Once again we assume that the U.S. takes full advantage of the available technology. In other words, every firm in the U.S. operates the best technology at the optimal scale, i.e., each of these firms operates technology 3 and has 1,600 units of capital. It is easily checked that the aggregate production function for the US implied by these assumptions if of the form $Y = AK^{0.4}L^{0.6}$. If India also operated at full efficiency (i.e.the production function is the same as in the US) we already know from our calculations based on the Lucas model that the ratio of output per worker in the two countries would be 6.35:1.

Our key assumption is that the distribution of firm sizes in India, by contrast with the US, includes a large fraction of firms that neither use the best technology nor operate at the optimal scale. The implicit premise is that in the U.S. there are enough people who are able and willing to invest 1,600 units if there is any money to be made, but this is not true in India because of borrowing constraints or other reasons.

A series of straightforward calculations gives the expression for the ratio of out per worker, which is also the ratio of wages in the two economies:

$$
\begin{aligned}
& (\frac{y_I}{y_U})^{\frac{1}{1-\gamma}} = (\frac{w_I}{w_U})^{\frac{1}{1-\gamma}} \\
= \quad & \frac{N_I}{N_U}\frac{L_U}{L_I} \times \\
& \frac{[\lambda_1(A_I)^{\frac{1}{1-\gamma}}(K_1-\underline{K_1})^{\frac{\alpha}{1-\gamma}} + \lambda_2(A_IA_2/A_1)^{\frac{1}{1-\gamma}}(K_2-\underline{K_2})^{\frac{\alpha}{1-\gamma}} + \lambda_3(A_IA_3/A_1)^{\frac{1}{1-\gamma}}(K_3-\underline{K_3})^{\frac{\alpha}{1-\gamma}}]}{(A_UA_3/A_1)^{\frac{1}{1-\gamma}}(K_3-\underline{K_3})^{\frac{\alpha}{1-\gamma}}},
\end{aligned}
$$

where N_I and N_U are the numbers of firms in India and the U.S., and A_I and A_U represent the base levels of TFP. The only reason that $A_I \neq A_U$ is, as before, that the human capital levels vary. We continue to assume that $A_U = 2A_I$. N_I/N_U can be computed from the fact that the total demand for capital from these firms must exhaust the supply of capital: i.e.,

$$\frac{K_I}{K_U} = \frac{N_I[\lambda_1K_1 + \lambda_2K_2 + \lambda_3K_3]}{N_UK_3},$$

which, given the assumed parameter values, implies that $N_I/N_U = 0.3$, which can then be used to calculate $\frac{y_I}{y_U}$, which turns out to be almost exactly 1/10, not too far from the 1/11 that we found in the data.

We can also derive, as before, what this model tells us about the marginal product of capital in the U.S. Using the expression derived in the previous sub-section, it is easily shown that the ratio of the marginal product of capital in the U.S. to that in the biggest and best Indian firms will be given by $(\frac{A^I}{A^U})^{\frac{1}{1-\gamma}}(w_U/w_I)^{\frac{\gamma}{1-\gamma}}$,[39] which turns out to be 6.45. Given that the biggest Indian firms have a marginal product of 100%, the average U.S. firms should have a marginal product of 100/6.45=15.5%. This is obviously higher than the average stock market return but hardly beyond the reasonable range.

[39]The fact that the biggest firms in India are the same size as any U.S. firm obviously simplifies the calculation.

• Distribution of firm sizes

The most obvious advantage of the fixed cost approach is that we do not obtain the unreasonably large gap in the marginal products of capital between small and large firms within the same economy becoming too large that came out of the previous model. This underscores the importance of using evidence on cross-sectional differences within an economy to assess the validity of alternative models.

Finally, the success of this model in explaining the productivity gap depends, as in the case of the previous model, heavily on the assumption that the U.S. has many more firms than India. However, while in that model we needed the U.S. to have 32 times as many firms as in India to fit the observed productivity gap, here we are doing very well with the U.S. having $3\frac{1}{3}$ times as many.

How reasonable is the assumption that the U.S. has more firms than India? This is not an easy question to answer, mainly because we have no clear sense of what should count as a firm: Both economies have enormous numbers of tiny firms that reflect what people do on the side. In India these "firms" are concentrated in a few sectors, such as retailing or the collection of leaves, wood or waste products, which require little or no skills and can be done on part-time basis. In the U.S., the equivalent would be the numerous ways in which you end up owning a small business for tax purposes, such as part-time consulting, renting out part of your home, part-time telemarketing, etc. It is not clear which of these should count as legitimate firms from the point of view of our model and which of these should not.

A way to restate the same point is that by focusing on the median industry in the ASI data, we have effectively ignored the industries (like the ones listed above) which attract all those in India who have nowhere better to go. While there are only a few such industries, they are enormous, and quite unlike the rest of the industries: Among the industries listed in the table above, cotton spinning is probably most like what one of these industries looks like, and it is apparent that it is quite different from the rest—there are many more tiny firms.

However this is not a problem for what we are doing here. Starting with the Indian firm size distribution assumed in the above exercise, we could simply add many more of the smallest firms to the Indian firms size distribution, till we get to the point where India has the same number of firms as the US. Since we have increased the number of firms in India by $3\frac{1}{3}$ times while keeping the number of large firms (firms with 1,600 units of capital) constant, the share of these firms goes down to 3% (from 10%). These two versions of the Indian economy are reasonably similar, because the smallest firms do not add up to a large amount of capital, but it is obvious that this economy will be less productive than the previous one (since inefficient small firms will have a larger share of total capital), and hence we will actually get somewhat closer to the 11:1 productivity ratio that we were shooting for.

• Why doesn't capital flow to India?

Finally we subject this model to an additional test: The fact that in our model there are firms in India with returns in the neighborhood of 100% would suggest that there are many unexploited opportunities. We have already argued that there are many reasons why a U.S. bank could not just lend to an Indian firm, and thereby benefit from these opportunities. Nor is it easy for an American to borrow money in the U.S. and set up a firm in India: Once he is in India he may be beyond the reach of U.S. law and for that reason alone, lenders will shy away from him. What is much more plausible, however, is that a U.S. entrepreneur moves to India to invest his money in these opportunities. The question is why this does not happen more often.

There are some obvious answers to this question: If the reason why these opportunities have not already been taken is the lack of secure property rights in India, there is no reason why foreigners would be particularly keen to invest in India. On the other hand, if the problem is that Indians do not have the capital or that they fear the risk exposure or that they are simply unaware of the opportunity, to take some plausible alternatives, a well-diversified wealthy U.S. investor may well be attracted to move to India and start a firm.

How much money would such an investor make? To answer this we start by observing from 7 that the production function in the largest Indian firms can be written as $C(K-800)^{1/2}$, where $C = A_3^{\frac{1}{1-\gamma}} \left[\frac{\gamma}{w}\right]^{\frac{\gamma}{1-\gamma}}$. Of this, a fraction 3/5 goes to wages. Profits are therefore given by $\frac{2}{5}C(K-800)^{1/2}$. Since this firm has 1,600 units of capital, and the marginal product of capital in this firm was assumed to be 100%, it follows that

$$\frac{1}{5}C(800)^{-1/2} = 1,$$

or

$$C = 141.42.$$

The opportunity cost of capital for a U.S. investor is 9%. The optimal investment in this Indian firm for a U.S. investor who can invest as much as he wants will be given by the solution to

$$(141.42)(0.2)(K-800)^{-1/2} = 0.09.$$

This tell us that the optimal investment is $K = 99564$. The total after-wage income generated by the firm is $(0.4)(141.42)(99564-800)^{1/2} = 17777$. This is in units of the smallest firm. We know that the biggest firms in our model are 1,600 times as large as the smallest firms and from the table above, such firms have Rs. 36 million worth of capital in the median industry. The smallest firm therefore has Rs.22,500 worth of capital, which implies that the U.S. investor will earn 17777(22500)=Rs. 400 million on his investment of (99564)(22500)=Rs. 2.24 billion. This is a net gain of about Rs. 200 million, or about 4 million dollars.

Is this a large enough gain to tempt someone to leave his home and family and settle in India? For someone with an average income, obviously. But no one with an average income has 50 million dollars of his own that he is willing to put into a single project in India. Anyone who is willing to do it has to be very rich indeed—he must have $50 million several times over. How many people that are so wealthy are willing to give up their life in the U.S. for an extra $4 million per year?

In other words, while the model developed in this section generates very large productivity losses, it does not offer any one person the possibility of arbitraging these unexploited opportunities to become enormously rich. This is because diminishing returns set in quite fast.

5.2.1 Taking Stock

We started by describing some of the major puzzles left unanswered by the neo-classical model, and in particular the productivity gap between rich and poor countries. The coexistence of high and low returns to investment opportunities, together with the low average marginal product of capital, suggested that some of the answer might lie in the misallocation of capital. The microeconomic evidence indeed suggests that there are some sources of misallocation of capital, including credit constraints, institutional failures, and others. In this section, we have seen that, combined with multiple technological options and a fixed cost of upgrading to better technologies, a model based on misallocation of capital does quite well in terms of explaining the productivity gap. The value of the marginal productivity of capital in the U.S. predicted by this model is only marginally too high, and the degree of variation in the marginal product of capital within a single economy that the model requires is not implausibly large.

Of course the model does make unrealistic assumptions—there is, for example, surely some amount of inefficiency in the U.S., and some U.S. firms are surely more productive than others. On the other hand, we have also ignored many reasons why Indian firms may be less efficient than they are in our model. For example, our current model assumes that only 10% of the firms, who use less than 1% of the capital stock and produce less than 1% of the output, use the least efficient technology whereas the MGI report on the apparel sector tells us that almost 55% of the output of the sector is produced by tailors who still use primitive technology. We also assumed that 10% of Indian firms are as productive as the best U.S. firms. Clearly that fraction could be smaller.

We also assumed that everyone is equally competent. In the real world, imperfect credit markets, for example, drives down the opportunity cost of capital and this encourages incompetent producers to stay in business. In the model, we assume that all large firms earn high returns but in reality there are probably some large firms that have much lower productivity (anywhere down to 9% per year would be consistent with our model). This too will drive down productivity. In a recent paper, Caselli and Gennaioli (2002) try to calibrate the impact of this factor in the context of a dynamic model with credit

constraints. They show that in steady state this can generate productivity losses of 20% or so. We will argue in the next section that this severely understates the potential productivity gap starting from an arbitrary allocation of capital.

6 Towards a Non-aggregative Growth Theory

6.1 An Illustration

The presumption of neo-classical growth theory was that being a citizen of a poor country gives one access to many exciting investment opportunities, which eventually lead on to convergence. The point of the previous section was to argue that most citizens of poor countries are not in a position to enjoy most of these opportunities, either because markets do not do what they ought to or the government does what it ought not to, or because people find it psychologically difficult to do what is expected of them.

What can we say about the long-run evolution of an economy where there are rewarding opportunities that are not necessarily exploited? In this section we will explore this question under the assumption that the only source of inefficiency in this economy comes from limited access to credit. The goal is to illustrate what non-aggregative growth theory might look like, rather than to suggest an alternative canonical model.

The model we have in mind is as follows: There are individual production functions associated with every participant in this economy that are assumed to be identical and a function of capital alone ($F(K)$) but otherwise quite general. In particular, we do assume that they are concave. Individuals maximize an intertemporal utility function of the form:

$$\begin{aligned} \sum_{t=0}^{\infty} \delta^t U(C_t), 0 &< \delta < 1 \\ U(C_t) &= \frac{c^{1-\phi}}{1-\phi}, \phi > 0. \end{aligned}$$

People are forward-looking and at each point of time they choose consumption and savings to maximize lifetime utility. However, the maximum amount they can borrow is linear and increasing in their wealth and decreasing in the current interest rate: An individual with wealth w can borrow up to $\lambda(r_t)w$. Credit comes from other members of the same economy and the interest rate clears the credit market. We do not assume that everyone starts with the same wealth, but rather that at each point of time there is a distribution of wealth that evolves over time.

This model is a straightforward generalization of the standard growth model. What it tells us about the evolution of the income distribution and efficiency depends, not surprisingly, on the shape of the production function.

The simplest case is that of constant returns in production. In this case, inequality remains unchanged over time, and production and investment is always efficient.

With diminishing returns, greater inequality can lead to less investment and less growth, because the production function is concave. However, inequality falls over time and in the long run no one is credit constrained, although we do not necessarily get full wealth convergence. The long run interest rate converges to its first best level, and hence investment is efficient. To see why this must be the case, note first that because of diminishing returns the poor always have more to gain from borrowing and investing than the rich. In other words, the rich must be lending to the poor. As long as the poor are credit constrained, they will earn higher returns on the marginal dollar than their lenders, i.e., the rich (that is what it means to be credit constrained). As a result, they will accumulate wealth faster than the rich and we will see convergence. This process will only stop when the poor are no longer credit constrained, i.e., they are rich enough to be able to invest as much as they want.

With increasing returns, inequality increases over time; we converge to a Gini coefficient of 1. Wealth becomes more and more concentrated with only the richest borrowing and investing. Because there are increasing returns, this is also the first best outcome. The logic of this result is very similar to the previous one: Now it is the rich who will be borrowing and the poor who will be lending, with the implication that the rich are the ones who are credit constrained and the ones earning high marginal returns. Therefore, they will accumulate wealth faster and wealth becomes increasingly concentrated.

Finally we consider the case of "S-shaped" production functions, which are production functions that are initially convex and then concave. The Cobb-Douglas with an initial set-up cost discussed at length in section 5.2 is a special case of this kind of technology.

What happens in the long run in this model depends on the initial distribution of income. When the distribution is such that most people in the economy can afford to invest in the concave part of the production function, the economy converges to a situation that is isomorphic to the diminishing returns case, with the entire population "escaping" the convex region of the production function.

The more unusual case is the one where some people start too poor to invest in the concave region of the production function. The poorer among such people will earn very low returns if they were to invest and therefore will prefer to be lenders. Now, as long as the interest rate on savings is less than $1/\delta$, they will decumulate capital (since the interest is less than the discount factor) and eventually their wealth will go to zero. On the other hand, anyone in this economy who started rich enough to want to borrow will stay rich, even though they are also dissaving, in part because at the same time they benefit from the low interest rates. The economy will converge to a steady state where the interest rate is $1/\delta$, those who started rich continue to be rich and those who started poor remain poor (in fact have zero wealth).

This is classic poverty trap: Moreover, since no one escapes from poverty, nor falls into it, there is a

continuum of such poverty traps in this model. This kind of multiplicity is, however, fragile with respect to the introduction of random shocks that allow some of the poor to escape poverty and impoverish some of the rich.

Even in a world with such shocks there can be more than one steady state: The reason is that the presence of lots of poor people drives down interest rates, and low interest rates make it harder for the poor to save up to escape poverty even with the help of a positive shock. As a result, in an economy that starts with lots of poor people, a greater fraction of people may remain poor.

The key to this multiplicity is the endogeneity of the interest rate. It is the pecuniary externality that the poor inflict on other poor people that sustains it. This is why such poverty traps are sometimes called *collective poverty traps*, in contrast to the *individual poverty traps* described above.

The investigation of the evolution of income distribution in models with credit constraints and endogenous interest rates goes back to Aghion and Bolton (1997), Matsuyama (2000, 2003) and Piketty (1997) emphasize the potential for collective poverty traps in a variant of this model, without the forward-looking savings decisions.

This class of models is a part of a broader group of models which study the simultaneous evolution of the occupational structure, factor prices and the wealth distribution in a model with credit constraints. Loury (1981) studied this class of models and showed that in the long run the neo-classical predictions tend to hold as long as the production function is concave. Dasgupta and Ray (1986) and Galor and Zeira (1993) provide examples of individual poverty traps in the presence of credit constraints and S-shaped production functions. Banerjee and Newman (1993) show the possibility of a collective poverty trap in a model with a S-shaped production function which is driven by the endogeneity of the wage–essentially high wages allow workers to become entrepreneurs easily, which keeps the demand for labor, and hence wages, high. Recent work by Buera (2003) shows that the multiplicity results in Banerjee and Newman survive in an environment where savings is based on expectations of future returns.[40] Ghatak, Morelli and Sjostrom (2001, 2002) and Mookerjee and Ray (2002, 2003) explore related but slightly different sources of individual and collective poverty traps.

6.2 Can we take this model to the data?

Models like the one we just developed (as well as political economy models that we do not discuss here[41]) have been invoked as motivation for a large empirical literature on the relationship between inequality

[40] On the possibility of collective poverty traps, see also Lloyd-Ellis and Bernhardt (2000), and Mookerjee and Ray (2002, 2003) .

[41] See Alesina and Rodrik (1994), Persson and Tabellini (1991) and Benhabib and Rustichini (1998). For a contrarian point of view, arguing that the premise of the political economy model argument does not hold true in the data, see Benabou (1996).

and growth in cross-country data. In 1996, Benabou cited 16 studies on the question, and the number has been growing rapidly since then, in part due to the availability of more complete data sets, due to the effort of Deininger and Squire (see Deininger and Squire (1996)), expanded by the World Institute for Development Economics Research (WIDER). However, it is not clear that if we were to take this class of models seriously, they would justify estimating relationships like the ones that are in the literature: First because the exact form of the predicted relationship between inequality and growth depends on the shape of the production function. Imposing the assumption that there are diminishing returns helps in this respect, but with this assumption functional form issues loom large. Finally, it is not clear how, given the model's structure, we can avoid running into serious identification problems.

In this section, we evaluate whether, given these concerns, estimating the relationship between inequality and growth in a cross-country data set remains useful. Having concluded that it has, at best, very limited use, we discuss an alternative approach based on calibrating non-aggregative models using micro data.

6.2.1 What are the empirical implications of the above model?

Functional Form Issues With constant returns to scale, distribution is irrelevant for growth. With diminishing returns, an exogenous mean-preserving spread in the wealth distribution in this economy will reduce future wealth and, by implication, the growth rate. However, the impact depends on the level of wealth in the economy: Once the economy is rich enough that everyone can afford the optimal level of investment, inequality should not matter. The estimated relationship between inequality and growth should therefore allow for an interaction term between inequality and mean income. Moreover, an economy closer to the steady state has both lower inequality and lower growth. This has two implications for the estimation of the inequality growth relationship. First, the fact that the economy becomes more equal as it grows tends to generate a spurious positive relation between growth and inequality, both in the cross-section as well as in time-series. As a result, both the cross-sectional and the first differenced (or fixed effects) estimates of the effect of inequality on growth run the risk of being biased upwards, compared to the true negative relation that we might have found if we had compared economies at the same mean wealth levels. Moreover, consider a variant of the model where there are occasional shocks that increase inequality. Since the natural tendency of the economy is towards convergence, we should expect to see two types of changes in inequality: Exogenous shocks that increase inequality and therefore reduce growth and endogenous reductions in inequality that are also associated with a fall in the growth rate. In other words, measured changes in inequality in either direction will be associated with a fall in growth.

Controlling properly for the effect of mean wealth (or mean income), is therefore vital for getting

meaningful results. The usual procedure is to control linearly (as in most other growth regressions) for the mean income level at the beginning of the period. It is, however, not clear that there is any good reason why the true effect should be linear. Moreover, it seems plausible that different economies will typically have different λs, and therefore will converge at different rates.

The model also tells us that while initial distribution matters for the growth rate, it only matters in the short run. Over a long enough period, two economies starting at the same mean wealth level will exhibit the same average growth rate. In other words, the length of the time period over which growth is measured will affect the strength of the relationship between inequality and growth.

The preceding discussion assumed that the interest rates converged. As we noted, that does not need to be the case. If we do not assume it, variants of the simple concave economy may no longer converge, even in the weaker sense of the long-run mean wealth being independent of the initial distribution of wealth. Intuitively, poor economies will tend to have high interest rates, and this in turn will make capital accumulation difficult (note that $\lambda' < 0$) and tend to keep the economy poor.[42] This effect reinforces the claim made above that inequality matters most in the poorest economies.[43] This economy can have a number of distinct steady states that are each locally isolated. This means that small changes in inequality can cause the economy to move towards a different and further away steady state, making it more likely that the relationship will be non-linear.

With increasing returns, growth rates increase with a mean preserving spread in income. As the economy grows, it also becomes more unequal. Interpreting the relationship between inequality and growth is difficult even after controlling for convergence.

In the S-shaped returns case, the relationship between inequality and growth can be negative or positive depending on the initial distribution, and the size of the increase. For example, if everybody is very poor (on the left of the convex zone), a small increase in inequality will reduce growth, but increasing inequality enough may push more people to the point where they are able to take advantage of the more efficient technology, and increases in inequality will increase growth. The relation between inequality and growth delivered by this model is clearly non-monotonic. Moreover, the strong convergence property does not hold in general. In other words, the growth rate of wealth may *jump up* once the economy is rich enough, with the obvious implication that economies with higher mean wealth will not necessarily grow more slowly. In other words, the effect of mean wealth, that is the so-called convergence effect, may not be monotonic in this economy. Linearly controlling for mean wealth therefore does not guarantee that

[42] See Piketty (1997). For a more general discussion of the issue of convergence in this class of models, see Banerjee and Newman (1993).

[43] There is, however, a counteracting effect: Poorer economies with high levels of inequality may actually have low interest rates because a few people may own more wealth than they can invest in their own firms, and the rest may be too poor to borrow. For a model where this effect plays an important role, see Aghion and Bolton (1997).

we will get the correct estimate of the effect of inequality. It is worth noting that this economy will have a connected continuum of steady states. This means that after a shock the economy will not typically return to the same steady state. However, since it does converge to a nearby steady state, this is not an additional source of non-linearity.

Identification Issues Even if we could agree on a specification that is worth estimating, it is not clear how we can use cross-country data to estimate it. Countries, like individuals, are different from each other. Even in a world of perfect capital markets, countries can have very different distributions of wealth because, for example, they have different distributions of ability. There is no causal effect of inequality on growth in this case, but they could be correlated for other reasons. For example, cultural structures (such as a caste system) may restrict occupational choices and therefore may not allow individuals to make proper use of their talents, causing both higher inequality and lower growth. Conversely, if countries use technologies that are differently intensive in skilled labor, those countries using the more skill intensive technology can have both more inequality and faster growth.

As we discussed in detail above, countries have different kinds of financial institutions, implying differences in the λ's in our model. Our basic model would predict that the country with the better capital markets is likely both to be more equal and to grow faster (at least once we control for the mean level of income). The correlation between inequality and growth will therefore be a downwards-biased estimate of the causal parameter, if the quality of financial institutions differs across countries.[44]

If these country specific effects were additive, one could control for them by including a country fixed-effect in the estimated relationship (or by estimating the model in first difference). This strategy will be valid only under the assumption that changes in inequality are unrelated to unobservable country characteristics that are correlated with changes in the growth rate. While this is a convenient assumption, it has no reason to hold in general. For example, skill-biased technological progress will lead both to a change in inequality and a change in growth rates, causing a spurious positive correlation between the two. To make matters worse, we have to recognize the fact that λ itself (and therefore the effect of inequality on growth at a given point in time) may be varying over time as a result of monetary policies or financial development, and may itself be endogenous to the growth process.[45]

The more general point that comes out of the discussion above is that unless we assume capital

[44] Allowing λ to vary also implies that the causal effects of inequality will vary with financial development (which is how Barro (2000) explains his results). The OLS coefficient is therefore a weighted average of different parameters, where the weights are the country-specific contributions to the overall variance in inequality (Krueger and Lindahl (2001)). It is not at all clear that we are particularly interested in this set of weights.

[45] See Acemoglu and Zilibotti (1997), and Greenwood and Jovanovic (1999), for theories of growth with endogenous financial development.

markets are extremely efficient (which, in any case, removes one of the important sources of the effect of inequality), changes in inequality will be partly endogenous and related to country characteristics which are themselves related to changes in the growth rate. Identifying the effect of inequality by including a country fixed-effect would not necessarily solve all the endogeneity problems. Moreover, as we discussed above, the theory suggests that the specification should allow for non-linear functional forms, and interaction effects, which will be difficult to accommodate with a fixed effect specification.

6.2.2 Empirical Evidence

The preceding discussion suggests that empirical exercises using aggregate, cross-country data to estimate the impact of inequality and growth will be extremely difficult to interpret. The results are also likely to be sensitive to the choice of specification. This may explain the variety of results present in the literature. A long literature (see Benabou (1996) for a survey) estimated a long run equation, with growth between 1990 and 1960 (say) regressed on income in 1960, a set of control variables, and inequality in 1960. Estimating these equations tended to generate negative coefficients for inequality. As the discussion in the previous subsection suggests, there are many reasons to think that this relationship may be biased upward or downwards. To address this problem, Li and Zou (1998) and Forbes (2000) used the Deininger and Squire data set to focus on the impact of inequality on short run (5 years) growth, and introduced a linear fixed effect.[46] The results change rather dramatically: The coefficient of inequality in this specification is positive, and significant. Finally, Barro (2000) used the same short frequency data (he is focusing on ten-year intervals), but does not introduce a fixed effect. He finds that inequality is negatively associated with growth in the poorer countries, and positively in rich countries.

Banerjee and Duflo (2003) investigate whether there is any reason to worry about the non-linearities that the theory suggests should be present. They find that when growth (or changes in growth) are regressed non-parametrically on changes in inequality, the relationship is an inverted U-shape. There is also a non-linear relationship between past inequality and the magnitudes of changes in inequality. Finally, there seems to be a negative relationship between growth rates and inequality lagged one period. These facts taken together, and in particular the non-linearities in these relationships (rather than the variation in samples or control variables), account for the different results obtained by different authors using different specifications.

Townsend and Ueda (2003) illustrate very clearly that this diversity of results is likely to come from the functional form and identification problems we just discussed. They simulate the 30 year evolution for 1000 economies based on a model similar to the ones we describe in this section, with non-linear

[46] Forbes (2000) also corrects for the bias introduced by introducing a lagged variable in a fixed effect specification by using the GMM estimator developed by Arellano and Bond (1991).

individual production function and credit constraints. The economies start in 1976, with a distribution of wealth calibrated to match the Thai economy in the same year. They then introduce aggregate and individual level shocks, and run regressions similar to the regressions run in the literature. Using the 1985 year as the "base year", they replicate the findings of the long run regressions. Using 1980 as the base year, they do not replicate those results. A regression similar to that of Forbes (2000) finds either positive or negative relationship, depending on sampling decisions. This exercise clearly show that aggregate cross country regressions are the wrong tool to evaluate the pertinence of this class of models.

6.3 Where do we go from here?

The discussion on functional form and identification, coupled with the empirical evidence of non-linearities even in very simple exercises, suggests that cross-country regressions are unlikely to be able to shed any meaningful light on the empirical relevance of models that integrate credit constraints and other imperfections of the credit markets. This is made worse by the poor quality of the aggregate data, despite the considerable efforts to produce consistent and reliable data sets. This contrasts with the increased availability of large, good quality, micro-economic data sets, which allow for testing specific hypotheses and derive credible identifying restrictions from theory and exogenous sources of variation. Throughout this chapter, we quoted many studies using micro-economic data which tested the micro-foundations for the models we discussed in this section.

Even a series of convincing micro-empirical studies will not be enough to give us an overall sense of how, together, they generate aggregate growth, the dynamics of income distribution, and the complex relationships between the two. The lessons of development economics will be lost to growth if they are not brought together in an aggregate context. In other words, it is not enough to use them to loosely motivate cross-sectional growth regression exercises—the discussion in this section is but an example of the misleading conclusions to which this can lead.

An alternative that seems likely to be much more fruitful is to try to build macroeconomic models that incorporate the features we discussed, and to use the results from the microeconomic studies as parameters in calibration exercises. The exercise we performed in section 5 of this chapter is an illustration of the kind of work that we can hope to do. There are a number of recent papers that in some ways go further in this direction than we have gone. In particular, Quadrini (1999) and Cagetti and Nardi (2003), for the U.S., and Paulson and Townsend (2004), for Thailand, try to calibrate a model with credit constraints to understand the correlation between wealth and the probability of becoming an entrepreneur. The paper by Buera (2003) mentioned above, emphasizes the fact that the long run correlation between wealth and entrepreneurship is weaker than the short run correlation, because as noted by Skiba (1978), Deaton (1992), Aiyagari (1994) and Carroll (1997), those who are credit constrained now but want to invest in

the future have a very strong incentive to save. This, Buera points out, reduces the ultimate efficiency cost of imperfect credit markets, though in spite of this, the person with the median ability level and the median starting wealth loses about 18% of lifetime welfare because of the credit constraints. Caselli and Gennaioli (2002) offer a slightly different calibration: Like Buera, they are worried about the fact that with credit constraints the biggest firms may not be run by the best entrepreneurs. This can be a source of very large productivity losses in the short run. However, since the best entrepreneurs will make the most money, in the long run their firms would necessarily become the largest, unless they died young. They show that even with this limiting factor, reasonable death rates would imply a 20% loss of productivity when we compare an economy without credit constraints with one that has them.

The calibrations so far have not attempted to see if the path of wealth distribution that results from calibrating this type of model matches the data. Our exercise above, for example, tries to match the distribution of firm sizes at a point of time, but says nothing about the path, while Buera does not try to match the data. The one exception is the papers by Robert Townsend and his collaborators based on Thai data (Jeong and Townsend (2003); Townsend and Ueda (2003)).

These papers, as well as those mentioned in the previous paragraphs, start from the assumption that every firm has a single, usually strictly concave, production technology. The only fixed cost comes from the fact that the firm needs an entrepreneur. As we saw above, this model does not do very well in terms of explaining the cross-sectional variation in the firm sector or the overall productivity gap, as compared to a model with a small number of alternative technologies and varying fixed costs. More generally, we need both a better empirical understanding of where the most important sources of inefficiency lie and better integration of this understanding when we assess the predictions of growth theory.

And perhaps above all, we need better growth theory: Our exercise at the beginning of this section was intended to advertise the possibility of a growth theory that does not assume aggregation. While we attempted to link the results to some relatively general properties of the production function, our analysis relies heavily on the fact that the inefficiency we assumed was in the credit market and that this took the form of a credit limit that was linear in wealth. One can easily imagine other ways for the credit market to be imperfect and other results from such models. Moreover, while the class of production technologies covered by our model was broader than usual, it does not include the (multiple-fixed-cost) technology that the previous section advocates.

There are, of course, other types of non-aggregative models: There are some examples of non-aggregative growth models that build on the inefficiency that comes from poorly functioning insurance markets.[47] There are also interesting attempts to build growth models that emphasize the fact that some

[47] See Banerjee and Newman (1991) for a theoretical model of non-aggregative growth based on imperfect insurance markets. Deaton and Paxson (1994) investigate some of empirical implications of this type of model using Taiwanese data.

people are favored by the government while others are not, and especially the fact that this changes over time in some predictable way (see Roland Benabou's contribution to this volume). Some interesting recent work has been done on the dynamic interplay between growth and political institutions (see the chapter by Acemoglu and Robinson in this volume) as well as between growth and social institutions (see Oded Galor's contribution to this volume, as well as Cole, Mailath and Postelwaithe (1992, 1998, 2001)). However, even more than in the case of the literature on credit markets and growth, it is not clear how much the insights from these models rely on specific details of how the environment or the imperfection was modeled and to what extent they can be seen as robust properties of this entire class of models.

There are also areas where growth theory has not really reached: We have no models that, for example, incorporate reputation-building or learning into growth theory. The same can be said about the entire class of behavioral models of underinvestment.

Finally, there is the open question of whether we gain anything by building grand models that incorporate all these different reasons for inefficiency in a single model. To answer this we would need to assess whether the fact that different forms of inefficiency interact with each other has empirically important consequences.

This is an exciting time to think about growth. We are beginning to see the contours of a new vision, both more rooted in evidence and more ambitious in its theorizing.

References

Acemoglu, Daron, and Fabrizio Zilibotti (1997) 'Was prometheus unbound by chance? Risk, diversification, and growth.' *Journal of Political Economy* 105(4), 709–51

—— (2001) 'Productivity differences.' *Quarterly Journal of Economics* 116(2), 563–606

Acemoglu, Daron, and Joshua Angrist (2001) 'How large are human-capital externalities? Evidence from compulsory schooling laws.' In *NBER Macroeconomics Annual 2000, Vol. 15,* ed. Ben Bernanke and Kenneth Rogoff (Cambridge and London: MIT Press) pp. 9–59

Acemoglu, Daron, Simon Johnson, and James Robinson (2001) 'The colonial origins of comparative development: An empirical investigation.' *American Economic Review* 91(5), 1369–1401

Aghion, Philippe, and Patrick Bolton (1997) 'A trickle-down theory of growth and development with debt overhang.' *Review of Economic Studies* 64(2), 151–72

Krussel and Smith (1998) and Angeletos and Calvet (2003) are attempts to calibrate the impact of imperfect insurance on welfare and growth.

Aghion, Philippe, and Peter Howitt (1992) 'A model of growth through creative destruction.' *Econometrica* 60(2), 323–351

Aghion, Philippe, Peter Howitt, and David Mayer-Foulkes (2004) 'The effect of financial development on convergence: Theory and evidence.' Working Paper 10358, National Bureau of Economic Research

Aiyagari, S. Rao (1994) 'Uninsured idiosyncratic risk and aggregate saving.' *Quarterly Journal of Economics* 109, 569–684

Aleem, Irfan (1990) 'Imperfect information, screening and the costs of informal lending: A study of a rural credit market in Pakistan.' *World Bank Economic Review* 3, 329–349

Alesina, Alberto, and Dani Rodrik (1994) 'Distributive politics and economic growth.' *Quarterly Journal of Economics* 109(2), 465–490

Anderson, Siwan, and Jean-Marie Baland (2002) 'The economics of roscas and intrahousehold resource allocation.' *Quarterly Journal of Economics* 117(3), 963–995

Angeletos, George-Marios, and Laurent Calvet (2003) 'Idiosyncratic production risk, growth and the business cycle.' Mimeo, MIT

Aportela, Fernando (1998) 'The effects of financial access on savings by low-income people.' Mimeo, MIT

Arellano, Manuel, and Stephen Bond (1991) 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations.' *Review of Economic Studies* 58(2), 277–297

Ashraf, Nava, Dean S. Karlan, and Wesley Yin (2004) 'Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines.' Mimeo, Harvard University

Banerjee, Abhijit V. (1992) 'A simple model of herd behavior.' *Quarterly Journal of Economics* 117(3), 797–817

—— (2003) 'Contracting constraints, credit markets and economic development.' In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. III,* ed. L. Hansen Dewatripont, M. and S. Turnovsky (Cambridge University Press)

—— (2004) 'Educational policy and the economics of the family.' *Journal of Development Economics* 74(1), 3–32

—— (2005) 'The two poverties.' In *Insurance Against Poverty,* ed. Stefan Dercan (Oxford University Press) pp. 59–75

Banerjee, Abhijit V., and Andrew Newman (1991) 'Risk bearing and the theory of income distribution.' *Review of Economic Studies* 58(2), 211–235

—— (1993) 'Occupational choice and the process of development.' *Journal of Political Economy* 101(2), 274–298

Banerjee, Abhijit V., and Esther Duflo (2000) 'Reputation effects and the limits of contracting: A study of the indian software industry.' *Quarterly Journal of Economics* 115(3), 989–1017

—— (2003) 'Inequality and growth: What can the data say?' *Journal of Economic Growth* 8, 267–299

—— (2004) 'Do firms want to borrow more? Testing credit constraints using a directed lending program.' Mimeo, MIT; also BREAD WP 2003-5, 2003

Banerjee, Abhijit V., and Kaivan Munshi (2004) 'How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur.' *Review of Economic Studies* 71(1), 19–42

Banerjee, Abhijit V., Esther Duflo, and Kaivan Munshi (2003) 'The (mis)allocation of capital.' *Journal of the European Economic Association* 1(2–3), 484–494

Banerjee, Abhijit V., Paul Gertler, and Maitreesh Ghatak (2002) 'Empowerment and efficiency: Tenancy reform in West Bengal.' *Journal of Political Economy* 110(2), 239–280

Bardhan, Pranab (1969) 'Equilibrium growth in a model with economic obsolescence of machines.' *Quarterly Journal of Economics* 83(2), 312–323

Bardhan, Pranab, and Christopher Udry (1999) *Development microeconomics* (Oxford, New York: Oxford University Press)

Barro, Robert J. (1974) 'Are government bonds net wealth?' *Journal of Political Economy* 82(6), 1095–1117

—— (1991) 'Economic growth in a cross section of countries.' *Quarterly Journal of Economics* 106(2), 407–443

—— (2000) 'Inequality and growth in a panel of countries.' *Journal of Economic Growth* 5(1), 5–32

Barro, Robert J., and Jong-Wha Lee (2000) 'International data on educational attainment updates and implications.' Working Paper 7911, National Bureau of Economic Research, September

Barro, Robert J., and Xavier Sala-I-Martin (1995) *Economic Growth* (New York: McGraw Hill)

Basta, S., Soekirman, D. Karyadi, and N. Scrimshaw (1979) 'Iron deficiency anemia and the productivity of adult males in Indonesia.' *American Journal of Clinical Nutrition* 32(4), 916–925

Becker, Gary (1981) *A Treatise on the Family* (Cambridge, MA: Harvard University Press)

Benabou, Roland (1996) 'Inequality and growth.' In *NBER Macroeconomics Annual 1996,* ed. Ben Bernanke and Julio J. Rotemberg (Cambridge and London: MIT Press) pp. 11–73

Benhabib, Jess, and A. Rustichini (1998) 'Social conflict and growth.' *Journal of Economic Growth* 1(1), 143–158

Benhabib, Jess, and Mark M. Spiegel (1994) 'The role of human capital in economic development: Evidence from aggregate cross-country data.' *Journal of Monetary Economics* 34(2), 143–174

Bennell, Paul (1996) 'Rates of return to education: Does the conventional pattern prevail in sub-saharan Africa?' *World Development* 24(1), 183–199

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, and Jonathan Zinman (2004) 'Pricing psychology.' Mimeo, Harvard University

Besley, Timothy (1995) 'Savings, credit and insurance.' In *Handbook of Development Economics,* ed. J. Behrman and T.N. Srinivasan, vol. 3A (Amsterdam: Elsevier Science) chapter 6, pp. 2123–2207

Besley, Timothy, and Anne Case (1994) 'Diffusion as a learning process:Evidence from HYV cotton.' Discussion Paper 174, RPDS, Princeton University

Bigsten, Arne, Anders Isaksson, Mans Soderbom, and Paul Collier Et Al (2000) 'Rates of return on physical and human capital in Africa's manufacturing sector.' *Economic Development and Cultural Change* 48(4), 801–827

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992) 'A theory of fads, fashion, custom, and cultural change as informational cascades.' *Journal of Political Economy* 100(5), 992–1026

Bils, Mark, and Peter Klenow (2000) 'Does schooling cause growth?' *American Economic Review* 90(5), 1160–1183

Binswanger, Hans, and Mark Rosenzweig (1986) 'Behavioural and material determinants of production relations in agriculture.' *Journal of Development Studies* 22(3), 503–539

Blanchflower, David, and Andrew Oswald (1998) 'What makes an entrepreneur?' *Journal of Labor Economics* 16(1), 26–60

Bottomley, A. (1963) 'The cost of administering private loans in underdeveloped rural areas.' *Oxford Economic Papers* 15(2), 154–163

Bourguignon, François, and Pierre-Andre Chiappori (1992) 'Collective models of household behavior.' *European Economic Review* 36, 355–364

Buera, Francisco (2003) 'A dynamic model of entrepreneurship with borrowing constraints.' Mimeo, University of Chicago

Cagetti, Marco, and M. De Nardi (2003) 'Entrepreneurship, frictions and wealth.' Staff Report 324, Federal Reserve Bank of Minneapolis

Carroll, Christopher (1997) 'Buffer-stock saving and the life cycle/permanent income hypothesis.' *Quarterly Journal of Economics* 112, 1–56

Carvalho, Irineu (2000) 'Household income as a determinant of child labor and school enrollment in Brazil: Evidence from a social security reform.' Mimeo, MIT

Case, Anne, and Christina Paxson (2001) 'Mothers and others: Who invests in children's health?' *Journal of Health Economics* 20(3), 301–328

Case, Anne, Christina Paxson, and Joseph Ableidinger (2002) 'Orphans in Africa.' Working Paper 9213, National Bureau of Economic Research

Case, I-Fen Lin Anne, and Sara McLanahan (2000) 'How hungry is the selfish gene?' *Economic Journal* 110(466), 781–804

Caselli, Francesco (2004) 'The missing input: Accounting for cross-country income differences.' Mimeo, Harvard University, forthcoming in 'Handbook of Development Economcis', Aghion, P. and S. Durlauf, eds.

Caselli, Francesco, and N. Gennaioli (2002) 'Dynastic management.' Mimeo, Harvard University

—— (2004) 'Dynastic management, legal reform and regulatory reform.' Mimeo, London School of Economics

Coate, Stephen, and Martin Ravallion (1993) 'Reciprococity without commitment: Characterization and performance of informal insurance arrangements.' *Journal of Development Economics* 40, 1–24

Cole, Harold, George Mailath, and Andrew Postlewaite (1992) 'Social norms, savings behavior and growth.' *Journal of Political Economy* 100(6), 1092–1126

—— (1998) 'Class systems and the enforcement of social norms.' *Journal of Public Economics* 70(1), 5–35

—— (2001) 'Investment and concern for relative position.' *Review of Economic Design* 6, 241–261

Commission on Macroeconomics, and Health (2001) *Macroeconomics and health: investing in health for economic development: Report* (Geneva: World Health Organization)

Conley, Tim, and Christopher Udry (2003) 'Learning about a new technology: Pineapple in Ghana.' Mimeo, Northwestern University

Dasgupta, A. (1989) *Reports on Informal Credit Markets in India: Summary* (New Delhi: National Institute of Public Finance and Policy)

Dasgupta, Partha, and Debraj Ray (1986) 'Inequality as a determinant of malnutrition and unemployment: Theory.' *The Economic Journal* 96(384), 1011–1034

Deaton, Angus (1992) *Understanding Consumption* (Oxford: Oxford University Press)

—— (1997) *The Analysis of Household Surveys* (World Bank, International Bank for Reconstruction and Development)

Deaton, Angus, and Christina Paxson (1994) 'Intertemporal choice and inequality.' *Journal of Political Economy* 1-2(3), 437–467

Deininger, Klaus, and Lyn Squire (1996) 'A new data set measuring income inequality.' *World Bank Economic Review* 10, 565–591

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer (2002) 'The regulation of entry.' *Quarterly Journal of Economics* 117(1), 1–37

—— (2003) 'Courts.' *Quarterly Journal of Economics* 118(2), 453–518

Do, Toan, and Laksmi Iyer (2003) 'Land rights and economic development: Evidence from Vietnam.' Mimeo, Harvard Business School

Dreze, Jean, and Geeta Gandhi Kingdon (2001) 'School participation in rural India.' *Review of Development Economics* 5(1), 1–24

Duflo, Esther (2001) 'The medium run effects of educational expansion: Evidence from a large school construction program in Indonesia.' Working Paper 2003-2, Bureau for Research in Economic Analysis of Development. forthcoming, Journal of Developemnt Economics

—— (2003) 'Poor but rational.' Mimeo, MIT

Duflo, Esther, Michael Kremer, and James Robinson (2003) 'Understanding technology adoption: Fertilizer in Western Kenya, preliminary results from field experiments.' Mimeo, MIT

Durlauf, Steven (1993) 'Nonergodic economic growth.' *Review of Economic Studies* 60(2), 349–366

Eaton, Jonathan, and Samuel Kortum (2001) 'Trade in capital goods.' *European Economic Review* 45(7), 1195–1235

Edmonds, Eric (2004) 'Does illiquidity alter child labor and schooling decisions? evidence from household responses to anticipated cash transfers in South Africa.' Working Paper 10265, National Bureau of Economic Research

Ellison, Glenn, and Drew Fudenberg (1993) 'Rules of thumbs for social learning.' *Journal of Political Economy* 101(4), 93–126

Ellison, Glenn, and Edward Glaeser (1997) 'Geographic concentration in U.S. manufacturing industries: A dartboard approach.' *Journal of Political Economy* 105(5), 889–927

—— (1999) 'Geographic concentration of industry: Does natural advantage explain agglomeration?' *American Economic Review* 89(2), 311–316

Eswaran, Mukesh, and Ashok Kotwal (1985) 'A theory of contractual structure in agriculture.' *American Economic Review* 75(3), 352–367

Fafchamps, Marcel (2000) 'Ethnicity and credit in African manufacturing.' *Journal of Development Economics* 61, 205–235

Fafchamps, Marcel, and Susan Lund (2003) 'Risk-sharing networks in rural Philippines.' *Review of Economic Studies* 71(2), 261–287

Fazzari, S., G. Hubbard, and B. Petersen (1988) 'Financing constraints and corporate investment.' *Brookings Papers on Economic Activity* 0(1), 141–195

Forbes, Kristin J. (2000) 'A reassessment of the relationship between inequality and growth.' *American Economic Review* 90(4), 869–887

Foster, Andrew D., and Mark R. Rosenzweig (1995) 'Learning by doing and learning from others: Human capital and technical change in agriculture.' *Journal of Political Economy* 103(6), 1176–1209

—— (1996) 'Technical change and human-capital returns and investments: Evidence from the green revolution.' *American Economic Review* 86(4), 931–953

—— (1999) 'Missing women, the marriage market and economic growth.' Mimeo, University of Pennsylvania

Foster, Andrew D., and Mark Rosenzweig (2000) 'Technological change and the distribution of schooling: Evidence from green revolution in India.' Mimeo, Brown University

Freeman, Richard (1986) 'Demand for education.' In *Handbook of Labor Economics,* ed. O. Ashenfelter and A. Layard, vol. 1 (Netherlands: Elsevier Publishers) chapter 6

Freeman, Richard, and Remco Oostendorp (2001) 'The occupational wages around the world data file.' *International Labour Review* 140(4), 379–401

Galor, Oded, and Joseph Zeira (1993) 'Income distribution and macroeconomics.' *Review of Economic Studies* 60(1), 35–52

Gelos, R. Gaston, and Alejandro Werner (2002) 'Financial liberalization, credit constraints, and collateral: Investment in the Mexican manufacturing sector.' *Journal of Development Economics* 67(1), 1–27

Gertler, Paul, and Jonathan Gruber (2002) 'Insuring consumption against illness.' *American Economic Review* 92(1), 51–76

Gertler, Paul J., and Simone Boyce (2002) 'An experiment in incentive-based welfare: The impact of PROGESA on health in Mexico.' Mimeo, University of California, Berkeley

Ghatak, Maitreesh, Massimo Morelli, and Tomas Sjostrom (2001) 'Occupational choice and dynamic incentives.' *Review of Economic Studies* 68(4), 781–810

Ghatak, Maitreesh, Massimo Morelli, and Tomas Sjostrom (2002) 'Credit rationing, wealth inequality, and allocation of talent.' Mimeo, London School of Economics

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

Goldstein, Markus, and Christopher Udry (1999) 'Agricultural innovation and resource management in ghana.' Mimeo, Yale University; Final Report to IFPRI under MP17

—— (2002) 'Gender, land rights and agriculture in Ghana.' Mimeo, Yale University

Greenwood, Jeremy, and Boyan Jovanovic (1999) 'The information technology revolution and the stock market.' *American Economic Review* 89(2), 116–22

Grossman, Gene, and Elhanan Helpman (1991) *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press)

Gugerty, Mary Kay (2000) 'You can't save alone: Testing theories of rotating savings and credit associations.' Mimeo, Harvard University

Hart, O., and J. Moore (1994) 'A theory of debt based on inalienability of human capital.' *Quarterly Journal of Economics* 109, 841–879

Heston, Alan, Robert Summers, and Bettina Aten (2002) 'Penn World Table version 6.1.' Technical Report, Center for International Comparisons at the University of Pennsylvania

Hoff, Karla, and Joseph Stiglitz (1998) 'Moneylenders and bankers: Price-increasing subsidies in a monopolistically competitive market.' *Journal of Development Economics* 55(2), 485–518

Howitt, P., and D. Mayer (2002) 'R&D, implementation and stagnation: A Schumpeterian theory of convergence clubs.' Working Paper 9104, National Bureau of Economic Research

Hsieh, Chang-Tai (1999) 'Productivity growth and factor prices in east asia.' *American Economic Review* 89(2), 133–138

Hsieh, Chang-Tai, and Peter J. Klenow (2003) 'Relative prices and relative prosperity.' Mimeo, University of California, Berkeley

Jacoby, Hanan (2002) 'Is there an intrahousehold flypaper effect? Evidence from a school feeding program.' *Economic Journal* 112(476), 196–221

Jeong, Hyeok, and Robert Townsend (2003) 'Growth and inequality: Model evaluation based on an estimation-calibration strategy.' Mimeo, University of Southern California

Johnson, Simon, John McMillan, and Christopher Woodruff (2002) 'Property rights and finance.' *American Economic Review* 92(5), 1335–1356

Jovanovic, Boyan, and Peter Rousseau (2003) 'Specific capital and the division of rents.' Mimeo, New York University

Jovanovic, Boyan, and Rafael Rob (1997) 'Solow vs. Solow: Machine prices and development.' Working Paper 5871, National Bureau of Economic Research

Kanbur, Ravi (1979) 'Of risk taking and the personal distribution of income.' *Journal of Political Economy* 87, 769–797

Karlan, Dean (2003) 'Using experimental economics to measure social capital and predict financial decisions.' Mimeo, Princeton University

Kihlstrom, R., and J. Laffont (1979) 'A general equilibrium entrepreneurial theory of firm formation based on risk aversion.' *Journal of Political Economy* 87, 719–748

Klenow, Peter J., and Andres Rodriguez-Clare (1997) 'The neoclassical revival in growth economics: Has it gone too far?' In *NBER Macroeconomics Annual 1997,* ed. Ben Bernanke and Julio J. Rotemberg (MIT Press)

Knack, Stephen, and Philip Keefer (1995) 'Institutions and economic performance: Cross-country tests using alternative institutional measures.' *Economics and Politics* 7(3), 207–227

Kremer, Michael, and Edward Miguel (2003) 'The illusion of sustainability.' Mimeo, University of California, Berkeley

Kremer, Michael, Sylvie Moulin, and Robert Namunyu (2003) 'Decentralization: A cautionary tale.' Mimeo, Harvard University

Krueger, Alan, and Mikael Lindahl (2001) 'Education for growth: Why and for whom?' *Journal of Economic Literature* 39(4), 1101–1136

Krueger, Anne (1967) 'Factor endowments and per capital income differences among countries.' *Economic Journal* 78, 641–659

Krussel, Per, and Anthony Smith (1998) 'Income and wealth heterogeneity in the macroeconomy.' *Journal of Political Economy* 106(5), 867–896

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny (1998) 'Law and finance.' *Journal of Political Economy* 106(6), 1113–1155

Laffont, Jean-Jacques, and Mohamed Salah Matoussi (1995) 'Moral hazard, financial constraints and sharecropping in El Oulja.' *Review of Economic Studies* 62(3), 381–399

Laibson, David (1991) 'Golden eggs and hyperbolic discounting.' *Quarterly Journal of Economics* 62, 443–477

Levine, Ross, and David Renelt (1992) 'A sensitivity analysis of cross-country growth regressions.' *American Economic Review* 82(4), 942–963

Li, Hongyi, and Heng-fu Zou (1998) 'Income inequality is not harmful for growth: Theory and evidence.' *Review of Development Economics* 2(3), 318–334

Li, R., X. Chen, H. Yan, P. Deurenberg, L. Garby, and J.G. Hautvast (1994) 'Functional consequences of iron supplementation in iron-deficient female cotton workers in Beijing China.' *American Journal of Clinical Nutrition* 59, 908–913

Lloyd-Ellis, H., and D. Bernhardt (2000) 'Enterprise, inequality and economic development.' *Review of Economic Studies* 67(1), 147–169

Long, Sharon K. (1991) 'Do the school nutrition programs supplement household food expenditures?' *Journal of Human Resources* 26, 654–678

Loury, Glenn C. (1981) 'Intergenerational transfers and the distribution of earnings.' *Econometrica* 49(4), 843–867

Lucas, Robert (1990) 'Why doesn't capital flow from rich to poor countries?' *American Economic Review* 80(2), 92–96

Lundberg, Shelly, and Robert Pollak (1994) 'Noncooperative bargaining models of marriage.' *American Economic Review* 84(2), 132–137

—— (1996) 'Bargaining and distribution in marriage.' *Journal of Economic Perspectives* 10(4), 139–158

Mankiw, Gregory, David Romer, and David Weil (1992) 'A contribution to the empirics of economic growth.' *Quarterly Journal of Economics* 107(2), 407–437

Manski, Charles (1993) 'Identification of exogenous social effects: The reflection problem.' *Review of Economic Studies* 60, 531–542

Matsuyama, Kiminori (2000) 'Endogenous inequality.' *Review of Economic Studies* 67(4), 743–759

—— (2003) 'On the rise and fall of class societies.' Mimeo, Northwestern University

Mauro, Paolo (1995) 'Corruption and growth.' *Quarterly Journal of Economics* 110(3), 681–712

McKenzie, David, and Christopher Woodruff (2003) 'Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises.' Working Paper 2003-20, Bureau for Research in Economic Analysis of Development

McKinsey Global Institute (2001) 'India: The growth imperative.' Report, McKinsey Global Institute

Miguel, Edward, and Michael Kremer (2003) 'Networks, social learning, and technology adoption: The case of deworming drugs in kenya.' Working Paper 61

—— (2004) 'Worms: Identifying impacts on education and health in the presence of treatment externalities.' *Econometrica* 72(1), 159–218

Mookherjee, Dilip, and Debraj Ray (2002) 'Contractual structure and wealth accumulation.' *American Economic Review* 92(4), 818–849

—— (2003) 'Persistent inequality.' *Review of Economic Studies* 70(2), 369–393

Morduch, Jonathan (1993) 'Risk production and saving: Theory and evidence from Indian households.' Mimeo, Harvard University

Munshi, Kaivan (2004) 'Social learning in a heterogeneous population: Technology diffusion in the Indian green revolution.' *Journal of Development Economics* 73(1), 185–213

Munshi, Kaivan, and Jacques Myaux (2002) 'Development as a process of social change: An application to the fertility transition.' Mimeo, Brown University

Munshi, Kaivan, and Mark Rosenzweig (2004) 'Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy.' Mimeo, Brown University

Murphy, Kevin, Andrei Shleifer, and Robert Vishny (1989) 'Industrialization and the big push.' *Journal of Political Economy* 97(5), 1003–1026

Murphy, Kevin, Andrei Shleifer, and Robert Vishny (1995) 'The allocation of talent: Implications for growth.' In *The economic analysis of rent seeking,* ed. Robert Tollison and Roger Congleton, vol. 49 of *Elgar Reference Collection: International Library of Critical Writings in Economics* (Ashgate; Elgar) pp. 301–328

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) pp. 657–68

Newman, Andrew (1995) 'Risk-bearing and 'Knightian' entrepreneurship.' Mimeo, Columbia University

Olley, G. Steven, and Ariel Pakes (1996) 'The dynamics of productivity in the telecommunications equipment industry.' *Econometrica* 64(6), 1263–1297

Parente, Stephen, and Edward Prescott (1994) 'Barriers to technology adoption and development.' *Journal of Political Economy* 102(2), 298–321

—— (2000) *Barriers to Riches: Walras-Pareto Lectures, vol. 3* (Cambridge, MA: MIT Press)

Paulson, Anna, and Robert Townsend (2004) 'Entrepreneurship and financial constraints in thailand.' *Journal of Corporate Finance* 10(2), 229–262

Persson, Torsten, and Guido Tabellini (1991) 'Is inequality harmful for growth? Theory and evidence.' *American Economic Review* 48, 600–621

Pessoa, Samuel, Pedro Cavalcanti-Ferreira, and Fernando Velloso (2004) 'The evolution of international output differences (1960-2000): From factors to productivity.' Mimeo, EPGE-FGV Rio

Piketty, Thomas (1997) 'The dynamics of the wealth distribution and the interest rate with credit rationing.' *The Review of Economic Studies* 64(2), 173–189

Powell, Christine, Sally Grantham-McGregor, and M. Elston (1983) 'An evaluation of giving the Jamaican government school meal to a class of children.' *Human Nutrition: Clinical Nutrition* 37C, 381–388

Powell, Christine, Sally Walker, Susan Chang, and Sally Grantham-McGregor (1998) 'Nutrition and education: A randomized trial of the effects of breakfast in rural primary school children.' *American Journal of Clinical Nutrition* 68, 873–879

Psacharopoulos, George (1973) *Returns to Education: An International Comparison* (San Francisco: Jossy Bass-Elsevier)

—— (1985) 'Returns to education: A further international update and implications.' *Journal of Human Resources* 20(4), 583–604

—— (1994) 'Returns to investments in education: a global update.' *World Development* 22(9), 1325–1343

—— (2002) 'Returns to investment in education: A global update.' *World Development* 22(9), 1325–1343

Qian, Nancy (2003) 'Missing women and the price of tea in china.' Mimeo, MIT

Quadrini, Vincenco (1999) 'The importance of entrepreneurship for wealth concentration and mobility.' *Review of Income and Wealth* 45, 1–19

Rauch, James (1993) 'Productivity gains from geographic concentration of human capital: Evidence from the cities.' *Journal of Urban Economics* 34(3), 380–400

Ravallion, Martin, and Shubham Chaudhuri (1997) 'Risk and insurance in village India: Comment.' *Econometrica* 65(1), 171–184

Ray, Debraj (1998) *Development Economics* (Princeton, N.J.: Princeton University Press)

—— (2003) 'Aspirations, poverty and economic change.' Mimeo, New York University

Restuccia, Diego, and Carlos Urrutia (2001) 'Relative prices and investment rates.' *Journal of Monetary Economics* 47(1), 93–121

Rosenstein-Rodan, Paul N. (1943) 'Problems of industrialization of eastern and south-eastern Europe.' *Economic Journal* 53, 202–211

Rosenzweig, Mark R., and Hans Binswanger (1993) 'Wealth, weather risk and the composition and profitability of agricultural investments.' *Economic Journal* 103(416), 56–78

Rosenzweig, Mark R., and Kenneth I. Wolpin (1993) 'Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India.' *Journal of Political Economy* 101(21), 223–244

Rosenzweig, Mark R., and T. Paul Schultz (1982) 'Market opportunities, genetic endowments, and intrafamily resource distribution: Child survival in rural India.' *American Economic Review* 72(4), 803–815

Sala-I-Martin, Xavier (1997) 'I just ran four million regressions.' Working Paper 6252, National Bureau of Economic Research

Schultz, T. Paul (2004) 'School subsidies for the poor: Eevaluating the Mexican PROGRESA poverty program.' *Journal of Development Economics* 74(1), 199–250

Shaban, Radwan (1987) 'Testing between competing models of sharecropping.' *Journal of Political Economy* 95(5), 893–920

Skiba, A. K. (1978) 'Optimal growth with a convex-concave production function.' *Econometrica* 46, 527–539

Stiglitz, J. (1969) 'The effects of income, wealth, and capital gains taxation on risk-taking.' *Quarterly Journal of Economics* 83(2), 263–283

—— (1974) 'Incentives and risk sharing in sharecropping.' *Review of Economic Studies* 41(2), 219–255

Strauss, John, and Duncan Thomas (1995) 'Human resources: Empirical modeling of household and family decisions.' In *Handbook of Development Economics,* ed. Jere Behrman and T.N. Srinivasan, vol. 3A (Amsterdam: North Holland) chapter 34, pp. 1885–2023

Strauss, John, and Duncan Thomas (1998) 'Health, nutrition, and economic development.' *Journal of Economic Literature* 36(2), 766–817

Svensson, Jakob (1998) 'Investment, property rights and political instability: Theory and evidence.' *European Economic Review* 42(7), 1317–1341

Thomas, Duncan (2001) 'Health, nutrition and economic prosperity: A microeconomic perspective.' Working Paper, Bulletin of the World Working Paper 7, Working Group 1, WHO Commission on Macroeconomics and Health

Thomas, Duncan, and Elizabeth Frankenberg (2002) 'Health, nutrition and prosperity: A microeconomic perspective.' *Bulletin of the World Health Organization* 80(2), 106–113

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, and Et Al (2003) 'Iron deficiency and the well being of older adults: Early results from a randomized nutrition intervention.' Mimeo, UCLA

Timberg, Thomas, and C. V. Aiyar (1984) 'Informal credit markets in India.' *Economic Development and Cultural Change* 33(1), 43–59

Tirole, Jean (1996) 'A theory of collective reputations (with applications to the persistance of corruption and to firm quality).' *Review of Economic Studies* 63(1), 1–22

Townsend, Robert (1994) 'Risk and insurance in village India.' *Econometrica* 62(4), 539–591

—— (1995) 'Financial systems in Northern Thai villages.' *Quarterly Journal of Economics* 110(4), 1011–1046

Townsend, Robert, and Kenichi Ueda (2003) 'Financial deepening, inequality, and growth: A model-based quantitative evaluation.' Mimeo, University of Chicago

Udry, Christopher (1990) 'Credit markets in Northern Nigeria: Credit as insurance in a rural economy.' *World Bank Economic Review* 4(3), 251–69

—— (1996) 'Gender, agricultural production, and the theory of the household.' *Journal of Political Economy* 101(5), 1010–1045

—— (2003) 'A note on the returns to capital in a developing country.' Mimeo, Yale University

United Nations Development Program (2001) *Human development report 2001: Making new technologies work for human development* (Oxford and New York: Oxford University Press)

Vermeersch, Christel (2002) 'School meals, educational achievement and school competition: Evidence from a randomized evaluation.' Mimeo, Harvard University

Young, Alwyn (1995) 'The tyranny of numbers: Confronting the statistical realities of the East Asian growth experience.' *Quarterly Journal of Economics* 110(3), 641–680

Zeldes, Stephen (1989) 'Consumption and liquidity constraints: An empirical investigation.' *Journal of Political Economy* 97(2), 305–346

Table 1: Returns to Education

Variable	Mincerian returns				log(teacher salary)	direct costs/benefits	total costs/benefits
Sample	Psacharopoulous	Psacharopoulous, extended	Psacharopoulos, high quality	Psacharopoulos, high quality			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	16.40	13.01	11.04	9.65	2.24	4.09	21.43
	(2.6)	(1.35)	(1.14)	(.46)	(.15)	(.21)	(1.63)
Mean years of	-0.72	-0.47	-0.27				
schooling	(.3)	(.16)	(.14)				
GDP/capita				-0.084		-0.034	-0.155
(*1000)				(.039)		(.019)	(.147)
lgdp					0.79		
					(.02)		
n	37	70	62	62	532	61	61
r^2	0.139	0.106	0.062	0.072	0.7902	0.05	0.018

Source: The data on returns to education was compiled starting from Psacharopoulous (2000) and extedended by surveying the literature. Appendix table 1 lists the data and the sources. The data on teacher salary is from Freeman and Oosterkberke. The data on pupil teacher ratio is from UNESCO (2000).

Appendix table 1: rate of returns to education and years of schooling.

Country	Continent	year	mincerian returns	years of schooling (Psacharopoulos)	Years of schooling (world bank)	Source	Data rating (bennel)	Additions to Psacharopoulos data
Argentina	South America	1989	10.3	9.1	8.83	Psacharopoulos (1994)		
Australia	Australia	1989	8	.	10.92	Cohn and Addison (1998)		
Austria	Europe	1993	7.2	.	8.35	Fersterer and Winter-Ebmer (1999)		
Bolivia	South America	1993	10.7	.	5.58	Patrinos (1995)		
Botswana	Africa	1979	19.1	3.3	6.28	Psacharopoulos (1994)	Poor	
Brazil	South America	1998	12.21	5.3	4.88	Verner (2001)		Added
Burkina Faso	Africa	1980	9.6	.	.	Psacharopoulos (1994)	Poor	
Cameroon*	Africa	1995	5.96		3.54	Appleton et al (1999)		Added
Canada	North America	1989	8.9	.	11.62	Cohn (1997)		
Chile	South America	1989	12	8.5	7.55	Psacharopoulos (1994)		
China	Asia	1993	12.2	.	6.36	Hossain (1997)		
Colombia	South America	1989	14	8.2	5.27	Psacharopoulos (1994)		
Costa Rica	South America	1992	8.50	.	6.05	Funkhouser (1998)		Added
Cote d'Ivoire	Africa	1987	13.10	6.9	.	Schultz (1994)	Poor	Added
Cyprus	Europe	1994	5.2	.	9.15	Menon (1995)		
Denmark	Europe	1990	4.5	.	9.66	Christensen and Westergard-Nielsen (1999)		
Dominican Re	South America	1989	9.4	8.8	4.93	Psacharopoulos (1994)		
Ecuador	South America	1987	11.8	9.6	6.41	Psacharopoulos (1994)		
Egypt	Africa	1997	7.80	.	5.51	Wahba (2000)		
El Salvador	South America	1992	7.6	.	5.15	Funkhouser (1996)		
Estonia	Europe	1994	5.4	10.9	.	Kroncke (1999)		
Ethiopia	Africa	1997	3.28	6	.	Krishnan, Selasie, Dercon (1989)	Poor	Added
Finland	Europe	1993	8.2	.	9.99	Asplund (1999)		
France	Europe	1977	10	6.2	7.86	Psacharopoulos (1994)		
Germany	Europe	1988	7.7	.	10.2	Cohn and Addison (1998)		
Ghana	Africa	1999	8.80	9.7	3.89	Frazer (1998)		Added
Greece	Europe	1993	7.6	.	8.67	Magoula and Psacharopoulos (1999)		
Guatemala	South America	1989	14.9	4.3	3.49	Psacharopoulos (1994)		
Honduras	South America	1991	9.3	.		Funkhouser (1996)		
Hong Kong	Asia	1981	6.1	9.1	4.8	Psacharopoulos (1994)		
Hungary	Europe	1987	4.3	11.3	9.13	Psacharopoulos (1994)		
India	Asia	1995	10.6	.	5.06	Kingdon (1998)		
Indonesia	Asia	1995	7	8	4.99	Duflo (2000)		
Iran	Asia	1975	11.6	.	5.31	Psacharopoulos (1994)	Poor	
Israel	Asia	1979	6.4	11.2	9.6	Psacharopoulos (1994)	Poor	
Italy	Europe	1987	2.7	.	7.18	Brunello, Comi and Lucifora (1999)		
Jamaica	South America	1989	28.8	7.2	5.26	Psacharopoulos (1994)	Poor	

Japan	Asia	1988	13.2	.	9.47	Cohn and Addison (1998)		
Kenya	Africa	1995	11.39	8	4.2	Appleton et al (1998)		Added
Korea	Asia	1986	13.5	8	10.84	Ryoo, Nam and Carnoy (1993)		
Kuwait	Asia	1983	4.5	8.9	7.05	Psacharopoulos (1994)	Poor	
Malaysia	Asia	1979	9.4	15.8	6.8	Psacharopoulos (1994)		
Mexico	South America	1997	35.31	.	7.23	Lopez-Acevedo (2001)		Added
Morocco	Africa	1970	15.8	2.9	.	Psacharopoulos (1994)	Poor	
Nepal	Asia	1999	9.7	3.9	2.43	Parajuli (1999)		
Netherlands	Europe	1994	6.4	.	9.36	Hartog, Odink and Smits (1999)		
Nicaragua	South America	1996	12.1	.	4.58	Belli and Ayadi (1998)		
Norway	Europe	1995	5.5	.	11.85	Barth and Roed (1999)		
Pakistan	Asia	1991	15.4	.	3.88	Katsis, Mattson and Psacharopoulos (1998)		
Panama	South America	1990	13.7	9.2	8.55	Psacharopoulos (1994)		
Paraguay	South America	1990	11.5	9.1	6.18	Psacharopoulos (1994)		
Peru	South America	1990	8.1	10.1	7.58	Psacharopoulos (1994)		
Philippines	South America	1998	12.6	8.8	8.21	Schady (2000)		
Poland	Europe	1996	7	.	9.84	Nesterova and Sabirianova (1998)		
Portugal	Europe	1991	8.6	.	5.87	Cohn and Addison (1998)		
Puerto Rico	South America	1989	15.1	.	.	Griffin and Cox Edwards (1993)		
Russian Federa	Europe	1996	7.2	11.7	.	Nesterova and Sabirianova (1998)		
Singapore	Asia	1998	13.1	9.5	7.05	Sakellariou (2001)		
South Africa*	Africa	1993	10.27	7.1	6.14	Mwabu and Schultz (1995)		Added
Spain	Europe	1991	7.2	.	7.28	Mora (1999)		
Sri Lanka	Asia	1981	7	4.5	6.87	Psacharopoulos (1994)		
Sudan	Africa	1989	9.3	10.2	2.14	Cohen and House (1994)		
Sweden	Europe	1991	5	.	11.41	Cohn and Addison (1998)		
Switzerland	Europe	1991	7.5	.	10.48	Weber and Wolter (1999)		
Taiwan	Asia	1998	19.01	9		Vere (2001)		Added
Tanzania*	Africa	1991	13.84	.	2.71	Mason and Kandker (1995)	Poor	Added
Thailand	Asia	1989	11.5	.	6.5	Patrinos (1995)		
Tunisia	Africa	1980	8	4.8	5.02	Psacharopoulos (1994)	Poor	
Uganda	Africa	1992	5.94		3.51	Appleton et al (1996)		Added
United Kingdo	Europe	1987	6.8	11.8	9.42	Psacharopoulos (1994)		
United States	North America	1995	10	.	12.05	Rouse (1999)		
Uruguay	South America	1989	9.7	9	7.56	Psacharopoulos (1994)		
Venezuela	South America	1992	9.4	.	6.64	Psacharopoulos and Mattson (1998)		
Vietnam	Asia	1992	4.8	7.9	.	Moock, Patrinos and Venkataraman (1998)		
Yugoslavia	Europe	1986	4.8	.	.	Bevc (1993)		

Zambia*	Africa	1995	10.65	5.46	Appleton et al (1999)	Added
Zimbabwe*	Africa	1994	5.57	5.35	Appleton et al (1999)	Added

Notes: This table updates Psacharopoulos (2002). The last column indicate which rate of returns were added by us
The data rating quality is from Bennell (1996), and concerns only African Countries