Why Does Misallocation Persist?*

Abhijit V. Banerjee
MIT
banerjee@mit.edu

Benjamin Moll
University of Chicago
bmoll@uchicago.edu

February 17, 2009

1 Introduction: Misallocation and Underdevelopment

There is growing interest in the view that underdevelopment may not just be a matter of lack of resources like capital, skilled labor, entrepreneurship or ideas but also a consequence of the misallocation or misuse of available resources. In particular Banerjee and Duflo (2005), Jeong and Townsend (2007), Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Bartelsman et al. (2008), Alfaro et al. (2008) and Buera et al. (2008) all argue that the extent of misallocation of resources in poor countries is large enough to explain a very large part of the TFP gap between rich and poor countries.

Evidence on the misallocation of resources takes many forms. There is evidence on interest rates suggesting that many smaller firms in developing countries borrow at interest rates of 50%, 80% or even higher.[1] This suggests that these firms must have marginal returns on capital that are even higher. These high estimates of the marginal returns are consistent with the direct evidence on the return on capital in small to medium sized firms in developing countries that we get from randomized experiments, natural experiments and other sources.[2]

However, we see a very different picture when we look at the evidence on the aggregate marginal product of capital for various developing countries. Caselli and Feyrer (2007) find

*We thank Daron Acemoglu, Olivier Blanchard, Steve Davis, Esther Duflo, Chang-Tai Hsieh and Pete Klenow for helpful comments and encouragement.

[1] Banerjee (2003) describes the evidence on this point and emphasizes that default is rare – so that these interest rates should be thought of as the rates that firms actually pay.

[2] See, for example, de Mel et al. (2008), Banerjee and Duflo (2008), Udry and Anagol (2006).

that the marginal product of capital is the same in poor and rich countries and is in fact below 10% everywhere. Swan (2008), using a different series for the prices of capital goods does find substantially higher estimates for developing countries, but even his high estimates are much lower than a lot of the firm level estimates for the marginal product of capital. Bai et al. (2006) come up with 20% as the aggregate marginal product of capital in China, down from 25% in the earlier period.

The obvious way to reconcile these two sets of facts is to assume that marginal products, contrary to what efficiency would require, are not equalized across firms – some firms have very high marginal products but a lot of other firms do not.

A second source of evidence involves fitting a production function to firm level data and directly estimating the distribution of marginal products or something related within an industry. Hsieh and Klenow (2009) estimate the distribution of TFPR, which turns out to be a geometric average of the marginal products of capital and labor and calculate that the ratios of 90th to 10th percentiles of TFPR are 5.0 in India, 4.9 in China and 3.3 in the U.S. Moreover the most productive firms (firms where the conventional measure of TFP is the highest) tend to be the most distorted in the direction of being too small in both countries, which amplifies the effect of the TFPR not being equalized.

A less structural version of the same exercise involves comparing the distribution of firm sizes across countries to argue that the distribution of firm sizes in most developing countries looks different from the presumed efficient US distribution (Alfaro et al., 2008). An alternative approach uses the correlation between firm size and the average product of labor as a measure of allocative efficiency, under the theory that the most productive firms should be the biggest (Bartelsman et al., 2008). Both these exercises yield some evidence that less developed countries have a joint distribution of firm size and productivity quite unlike the US.

Finally one could do a pure calibration exercise using some plausible parameter values and a model to put some magnitudes on the potential extent of output loss due to misallocation (as in Jeong and Townsend (2007), Restuccia and Rogerson (2008), Buera et al. (2008), Banerjee and Duflo (2005)). All of these papers find that 50% or more of the difference between rich and poor countries (or in the case of Jeong and Townsend, 73% of the increase in TFP in Thailand can be explained by the effects of misallocation under reasonable assumptions about parameter values

While each of these pieces of evidence has its limitation, taken together they strongly suggest that misallocation is quantitatively important as an empirical phenomenon.

2 Theorizing Misallocation

One very natural explanation of why there is so much misallocation, especially given the evidence presented above about the high rates of interest, is to blame asset markets: the inefficiency in the functioning of asset markets makes it harder for successful firms to acquire the assets they need to expand and simultaneously allows failed firms to survive (because the alternative of downsizing and putting the rest of the money in asset markets is unattractive). As a result high productivity firms underinvest in what they need – be it management ideas, new technology, marketing advice, reputation building or just new plants or machinery.

Focusing on asset markets would also be consistent with Hsieh and Klenow's result that most of the gain in both India and China would come from reallocating capital across firms. From the point of view of reallocating resources the key asset markets are the markets for land, financing and opportunities for risk diversification (which we will call risk capital). Of the primary assets of a firm, land (and what gets built on it) is the one where the physical adjustment costs are high everywhere in world. However while the acquisition of land has been a major issue in both India and China, this is less a problem at the firm level (unless it is a very large firm) than at the regional level, whereas the reallocation that Hsieh and Klenow are emphasizing is mostly between firms within the same region. On the other hand, there may well be constraints on selling land (though this seems unlikely since both India and China had a boom in residential real estate in this period, which made it extremely lucrative to sell existing land-holdings which were often in prime locations) and getting building permits and infrastructure connections are quite likely to have been a problem. At this stage we know too little, descriptively, about the workings of the urban "land" market in developing countries to say anything useful about this.

Finance and risk capital are of course the other two key assets that any firm needs (and the availability of which constrains its ability to then acquire machines, ideas, consulting, etc.) and there we know that the US financial infrastucture is much better. The banking sector in both India and China continues to be dominated by slow-moving and badly managed public sector banks and the system as a whole is notoriously ineffective in the enforcement of credit contracts, so that even the private sector is often unwilling to lend. The stock markets are not known for their effective regulations (in India things are said to have improved a lot, but only after 2000, while the data ends in 1995). And venture capital as an institutional form is more or less in its infancy in both countries.

However an alternative to acquiring these assets on the market is to accumulate them. The high rates of return faced by firms that are underinvesting, create a strong pressure for accumulation. Similarly, the lack of risk capital generates a precautionary savings motive which may drive firms to accumulate so much capital that they can self-insure. Both these forces generate forces towards eliminating the distortions across firms.

This does not mean, as we will emphasize later in the paper, that these asset market failures do not have aggregate consequences. But it does pose a challenge to explaining why distortions would not go away on their own; since underdevelopment is a persistent phenomenon, we need a theory that explains the persistence of misallocation.[3]

The next section sets up a simple model which helps us understand this challenge. In the interest of simplicity we suppress the risk capital issue by assuming away all risk. We come back to this issue in the closing section.

3 A Simple Model of Capital Accumulation with Credit Constraints

3.1 Preferences and Technology

Time is discrete. There is a continuum of households that are indexed by their ability z and their wealth a. At each point in time t, the state of the economy is some joint distribution $G_t(a, z)$. The marginal distribution of ability is denoted by $\mu(z)$ and its support by $[\underline{z}, \overline{z}]$.

Agents have preferences

$$\sum_{t=0}^{\infty} \beta^t u(c_t), \tag{1}$$

where u is strictly increasing, strictly concave and satisfies standard Inada conditions. Each

[3]There is some evidence on the rate of change in the extent of misalloction: Hsieh and Klenow (2009) report that for China, hypothetically moving to "U.S. efficiency" might have boosted TFP by 50% in 1998, and 30% in 2005. But for India, hypothetically moving to U.S. efficiency might have raised TFP around 40% in 1987 or 1991, and 59% in 1994, notwithstanding the fact that in this period India liberalized substantially. However there is ome danger of overinterpreting this evidence, since taking short term changes seriously asks a lot of the data.

household owns a private firm which uses k_t units of capital to produce $zf(k_t)$ units of output. We assume that the function f is strictly increasing but not necessarily concave. Capital depreciates at the rate δ.

3.2 Market Structure and Equilibrium

Denote by a_t an agent's wealth and by r_t the (endogenous) interest rate. Agents can rent capital k_t in a rental market at a rental rate $R_t = r_t + \delta$.[4] Then an agent's wealth evolves according to

$$a_{t+1} = zf(k_t) - (r_t + \delta)k_t + (1 + r_t)a_t - c_t \tag{2}$$

Agents face borrowing constraints.

$$k_t \leq \lambda(r_t)a_t \tag{3}$$

where $\lambda(\cdot)$ is continuous and non-increasing.[5] The production and savings/consumption decisions separate in a convenient way. Define the profit function

$$\pi_t(a, z) = \max_k \left\{ zf(k) - (r_t + \delta)k + (1 + r_t)a \quad \text{s.t.} \quad k \leq \lambda(r_t)a \right\}. \tag{4}$$

It is easy to see that this profit function is increasing in both its arguments. Also denote the optimal capital choice from this profit maximization problem by

$$k_t(a, z) = \min\{\lambda(r_t)a, k^u(z, r_t)\}$$

where

$$k^u(z, r_t) \equiv \arg\max_k \{zf(k) - (r_t + \delta)k\} \tag{5}$$

is unconstrained capital demand.

[4] Here the capital is accumulated by some intermediary who then rents it out to entrepreneurs. That the rental rate equals $r_t + \delta$ can be shown by a standard arbitrage argument. This way of stating the problem avoids carrying k_t as a state variable in the agent's problem.

[5] For example, suppose that agents can avoid the payment for rented capital $(r + \delta)k$ by incurring a cost proportional to capital usage, ϕk where we assume $\phi < r + \delta$. If they default they also lose their savings $(1 + r)a$. The enforcement constraint is $zf(k) - (r + \delta)k + (1 + r)a \geq zf(k) - \phi k$ so that $\lambda(r) = (1 + r)/(r + \delta - \phi)$. Note that, with other forms of credit market imperfections λ will generally also depend on ability z.

Summarizing, at each point in time t, each household solves

$$
\begin{aligned}
v_t(a,z) &= \max_{\{c_s,a_s\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \beta^{s-t} u(c_s) \quad \text{s.t.} \\
a_{s+1} &= \pi_s(a_s,z) - c_s, \forall s \geq t, \quad a_t = a.
\end{aligned}
\tag{6}
$$

The problem for each household can be written in recursive form:

$$
v_t(a,z) = \max_{a'} \; u[\pi_t(a,z) - a'] + \beta v_{t+1}(a',z). \tag{7}
$$

Note that the value function is indexed by t. This is because r_t varies over time, albeit exogenously from the point of view of the household. Denote the optimal choice of savings a' by $s_t(a,z)$. This is the policy function of a household with assets a and productivity z.

This paper studies capital misallocation. We find it useful to distinguish between two forms of misallocation.

Definition 1 *(i) We say there is capital misallocation on the* intensive *margin at time t if marginal products of capital $zf'(k_t(a,z))$ are not equalized across all agents who have positive levels of capital usage, $k_t(a,z) > 0$.*

(ii) We say there is capital misallocation on the extensive *margin if it is possible to redistribute capital from one agent to another individual with either an equal marginal product or zero capital and raise sum of their outputs.*

Misallocation on the intensive margin is misallocation in the conventional sense and is what the empirical evidence has been principally about(Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008). On the other hand, misallocation at the extensive margin cannot exist if there are no non-convexities in production and would not be picked up at all by the current methodologies for measuring misallocation, which focuses on the equalization of the marginal products. Therefore there may be much more misallocation than the data on marginal products suggests – in particular because there are talented people who never have enough money to set up a business and therefore we do not even see them in the data. Jeong and Townsend (2007) and Buera et al. (2008) attempt to get at this by making assumptions about the underlying distribution of talent.

A *competitive equilibrium* in this economy is defined in the usual way. That is, (i) individuals solve (6) taking as given the equilibrium time path for the interest rate $\{r_t\}_{t=0}^{\infty}$, and (ii) the capital market (which is the only market in this economy) clears at every point in time

$$\int k_t(a,z)G_t(da,dz) = \int aG_t(da,dz), \quad \text{all } t \geq 0. \tag{8}$$

The main question that we are interested in is whether misallocation disappears over time. The answer turns out to depend on the shape of the production function (in particular whether it exhibits diminishing returns or not) and on whether we are looking for misallocation at the intensive or the extensive margin – in particular, misallocation at the intensive margin tends to disappear in the long run even when the misallocation at the extensive margin does not. This is what we now turn to.

3.3 Diminishing Returns

In this sub-section, we make the following assumption

Assumption 1 *The function $f(k)$ is concave and satisfies standard Inada conditions.*

As already noted, there is no role for misallocation on the extensive margin with diminishing returns. The Euler equation corresponding to (6) is

$$u'(c_t) = \beta u'(c_{t+1})\left[1 + r_{t+1} + \lambda(r_{t+1})\psi_{t+1}\right] \tag{9}$$

where

$$\psi_t = \max\{zf'(\lambda(r_t)a_t) - (r_t + \delta), 0\}$$

is the Lagrange multiplier on the borrowing constraint (2). If there is no misallocation on the intensive margin at time t, marginal products are equalized across individuals and all multipliers are zero. In this case the unconstrained Euler equation

$$u'(c_t) = \beta u'(c_{t+1})(1 + r_{t+1}) \tag{10}$$

holds.

The following Lemma about the optimal savings policy function $s_t(a, z)$ is an adaptation of a result by Dechert and Nishimura (1983) and will be useful below. Note that it applies regardless of the assumptions on technologies f, for instance also for the case of (local) increasing returns.

Lemma 1 *The policy function $s_t(\cdot, z)$ is strictly increasing for all z.*

(All proofs appear in the Appendix.) This Lemma implies

Corollary 1 *Consider individuals with the same ability z. Their wealth trajectories never intersect, that is if $a_0 > \hat{a}_0$ then $a_t > \hat{a}_t$ for all t.*

We are now in the position to prove our main result about the asymptotic behavior of misallocation with diminishing returns.

Proposition 1 *Under Assumption 1 there is no misallocation asymptotically, that is* (10) *holds for all agents as $t \to \infty$.*

Informal versions of this claim have appeared in the literature in the past (for example in Banerjee and Duflo (2005)) and it is entirely intuitive. Indeed it does not depend on the particular form of credit constraints (3) we assume. For example, Moll (2009) analyzes a similar environment with the main exception that credit constraints take the form of endogenous and forward-looking limited enforcement constraints. Misallocation also disappears with this form of credit constraints. This makes clear that the logic behind the proof of Proposition 1 is very general. Note also what this proposition does and does not say. It says that misallocation disappears asymptotically as time $t \to \infty$. It does not say that this will happen in finite time. In fact, Moll (2009) finds that credit constraints always bind in finite time.

More can be said about steady states of this economy.

Definition 2 *A* steady state *is a competitive equilibrium with a constant interest rate $r_t = r^*$ and a constant aggregate capital stock $K_t = K^*$ for all t.*[6]

In line with Proposition 1, we can show that there will be no capital misallocation in steady state. Furthermore the interest rate equals the rate of time preference, $r^* = \rho$ where ρ is defined by $\beta \equiv (1+\rho)^{-1}$. First note that an interest rate greater than the rate of time preference,

[6]The definition of the aggregate capital stock is the obvious one, $K_t \equiv \int aG_t(da, dz)$.

$r^* > \rho$, is inconsistent with a steady state: from the Euler equation this would imply positive consumption growth $c_{t+1} > c_t$ for all agents which can only be true if the aggregate capital stock grows. We can also rule out an interest rate $r^* < \rho$. In the absence of credit constraints agents would dissave until they reach zero wealth. With credit constraints, this is not true anymore; instead individuals dissave until their wealth reaches a level satisfying

$$1 + \rho = 1 + r^* + \lambda(r^*)\psi^*(a, z) = 1 + r^* + \lambda(r^*)[zf'(\lambda(r^*)a) - r^* - \delta]. \tag{11}$$

That is, unconstrained agents only stop dissaving once they become constrained. But there cannot be only constrained agents in equilibrium; there have to be some lenders as well. This tells us that the interest rate must equal the rate of time preference.

Given this, no one is decumulating capital in the steady state. Now if some individuals were credit constrained in steady state they would have an incentive to accumulate wealth because the right-hand side of (11) would be greater than the left-hand side. Therefore the capital stock must be going up, contradicting the definition of a steady state. Summarizing, in any steady state of the economy with diminishing returns all agents must be unconstrained and the interest rate equals the rate of time preference.

The allocation of capital is then first-best and can be described by an aggregate production function. Individual capital usage $k^*(a, z)$ is therefore simply the same as in the standard neoclassical growth model

$$zf'(k^*(a, z)) = \rho + \delta, \quad \text{all } (a, z).$$

Recalling the definition of unconstrained capital demand $k^u(z, r)$ in (5), this implies that individual steady state capital stocks $k^*(a, z)$ are equal to the unconstrained capital demands $k^u(z, \rho)$. The unique aggregate steady state capital stock is first-best and equals

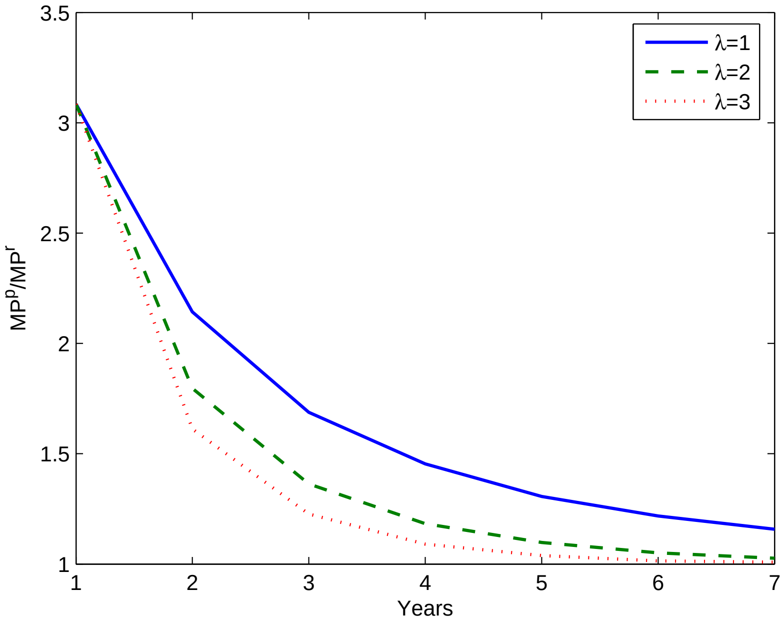
$$K^* = \int k^u(z, \rho)\mu(z)dz.$$

We have shown above that misallocation disappears asymptotically as $t \to \infty$. A natural question is: how long does this convergence of marginal products take? To get a feel for the length of time involved, we have conducted some numerical experiments.[7] We compare two

[7] Production takes the Cobb-Douglas form $f(k) = k^\alpha$. Utility is of the CRRA form with parameter σ. Borrowing constraints just take the simple form $k_t \leq \lambda a_t$ where $\lambda \geq 1$ is a constant. We choose the following

individuals, one rich and one poor. The borrowing constraint of the poor agent binds at time zero so much so that the poor agent has a *three* times higher marginal product than the rich agent, $z^p f'(k^p)/z^r f'(k^r) = 3$. Figure 1 plots the ratio of the marginal products over time for different λ. The speed at which the gap between the two marginal products narrows is

Figure 1: Convergence of Marginal Products



quite striking: Even if credit markets are completely shut down, $\lambda = 1$ the gap has almost disappeared after seven years. For better functioning credit markets convergence is even faster. With $\lambda = 3$ for example, the gap has essentially disappeared after five years. From a theoretical perspective, the result becomes less surprising if one keeps in mind that the model is a variant of the standard neoclassical growth model. With no credit markets, $\lambda = 1$, the models are in fact identical. It is well known that the speed of convergence of the neoclassical growth model is relatively high (King and Rebelo, 1993). With better working credit markets, this speed is only increased.

3.4 Local Increasing Returns

The assumption of global diminishing returns is not always a great description of how actual firms function – there is often a set-up cost involved in starting a business and more generally, non-convexities arise naturally from the fact that machines come in fixed sizes and ideas tend

very conventional parameter values: $\alpha = 0.3, \delta = 0.05, \beta = 0.95, \sigma = 2$.

to be indivisible. Following Skiba (1978) and Dechert and Nishimura (1983) we now allow for increasing returns over some range

Assumption 2 *$f(k)$ is convex over some range of k, but there is a $\hat{k} < \infty$ such that $f(k)$ is concave for $k > \hat{k}$.*

$f(k) = (k - \underline{k})^{\alpha}$, for $k \geq k$ and zero otherwise, is an example of this kind of production function.

In this section, we only look at steady states (as in Def. 2) – we are not able to guarantee that the economy converges to a steady state, though in our simulations it always does. As a result of the steady state assumption, in contrast with the earlier analysis, the problem of an agent is now stationary. Nevertheless, because the production function is non-convex, so is the agent's maximization problem. The result is that people at different wealth levels may exhibit radically different behaviors – those who are not too far below a particular non-convexity will save up and "cross" the non-convex region to get to the high returns available at high levels of investment, while those with only slightly less assets will prefer to dissave because the climb to get to the high returns is too far for them. Therefore where you converge to depends on where you started and there are multiple individual steady state wealth levels with different levels of output associated with them (these results are not formally demonstrated here but follow directly from the logic of such problems spelt out in Skiba (1978) and Dechert and Nishimura (1983)).

Note, however, that regardless of any non-convexities, it is still true that credit constrainted agents have a higher intertemporal marginal rate of substitution than unconstrained agents. This allows us to prove

Proposition 2 *Consider a steady state with constant interest rate r and constant aggregate capital stock K^* (definition 2). As $t \to \infty$, each individual is in steady state, there is no capital misallocation on the intensive margin, and $r = \rho$.*

While not reported here, we also carried out numerical exercises parallel to the ones reported above for the diminishing returns case, under the current assumption about the production function. Once again, except for those who are trying to "cross the non-convex region", convergence to an unconstrained state is relatively quick, which should surprise us since they essentially operate in a diminishing returns environment.[8]

3.5 Implications of These Results

These results tell us that unless convergence to an aggregate steady state fails (which we have not ruled out for the local returns increasing case) and the interest continues to fluctuate substantially even in the long run, we would expect to see misallocation at the intensive margin to dissappear relatively quickly. That does not have to mean that there is a no misallocation in the economy. Indeed with local increasing returns we can construct examples where the extensive margin misallocation is so large that steady state output as a ratio of first best steady state output is arbitarily close to zero.[9] But it does raise the question: Why do we see so much intensive margin misallocation in the data (remember all the misallocation that Hsieh and Klenow, for example, find, is at the intensive margin)?

4 Conclusion: Towards an Understanding of Persistence

Persistence is easy to explain if we are prepared to assume, as Restuccia and Rogerson (2008) do (for illustrative purposes) that there are ”taxes” on the firms that are permanently fixed. However since Hsieh and Klenow in particular only compare firms within an industry (and indeed get very similar results from comparing firms within the same industry within the same region) we need to explain why these taxes vary at the firm level. Moreover as, Restuccia and Rogerson, point out, to get large effects, the taxes need to be strongly positively correlated with firm level TFP.

The problem is explaining why there would be such large firm specific taxes. Most firms that are in the Hsieh and Klenow study are neither large enough (the median firm in the top quartile of the distribution has around two hundred employees) nor in another way so special (there are thousands of such firms) to attract special attention, one way or the other, from the political system in either of the two enormous countries. And while one could imagine firms

[8]Such non-convex problems are not much harder to solve computationally than their convex counterparts. This is because standard dynamic programming does not require the assumption of a concave period return function. This fact is, for example, used extensively in Dechert and Nishimura (1983).

[9]There is a long tradition of research on the effects of fixed costs combined with credit constraints on the long run performance of the economy (Galor and Zeira, 1993; Banerjee and Newman, 1993; Aghion and Bolton, 1997; Piketty, 1997).

getting into an especially friendly or unfriendly relationship with the local political bosses, it is not clear why the firms that are suffering could not move to a different area, or why someone without the political baggage could not buy out the firm.

It is true that India (but as far as we know, not China) has some policies - labor laws in particular - that specifically discriminate against larger firms. However Hsieh and Klenow (2009) report that their estimates of the potential productivity gain change by very little when they only equalize TFPR between firms within the same size quartile, and moreover, states in India with better labor laws do not do better in terms of TFPR dispersion.

The alternative approach to persistence relies on shocks. Within our framework there can be shocks to both assets (a) and ability(z) but any temporary shock to the profitability of the firm that occurs after investment has been chosen is isomorphic to an a shock, while any shock that affects the marginal product of investment that is yet to be carried out is like a z shock.

From the point of view of explaining the persistence of misallocation, shocks are important because they move people away from their steady states. Thus someone who used to have low z but now discovers that she has a high z will be massively under-invested and will show a high marginal product. The same will be true if a firm loses a large part of its capital stock.

Shocks however have another effect. As was pointed by Aiyagari (1994), the presence of shocks also changes savings behavior through the precautionary motive. In principle this additional accumulation might counteract the effect of the shocks. However as Angeletos and Calvet (2006) and Angeletos (2007) point out, this depends on exactly how the shocks are modeled. If shocks have an effect on the marginal product of capital (as our z shocks do), then they might actually discourage saving. Indeed just the savings discouragement effect, as Angeletos and Calvet calculate, can reduce the steady state capital stock by as much as 60%. However in their model no one is credit constrained and there is no misallocation at either margin.

In some ongoing work, Buera and Moll (2009) look at the effects of z shocks in a model with diminishing returns and credit constraints (where the precautionary savings effect is deliberately suppressed) and find that it is possible to find stationary states where the wedges are large and the productivity loss is substantial despite the strong tendency towards convergence that one finds in a model with diminishing returns.[10]

[10] In the model with a locally convex production technology, shocks can also knock individual firms below some threshold or push them above some threshold so that they converge towards a different steady state. This

Of course this is only possible because they assume shocks that are both frequent and relatively large. How plausible are such shocks?

Caselli and Gennaioli (2005) is one paper that takes a clear stand on this issue. They look at a model where there are only z shocks and these occur once in every generation, because the current owner of the business dies and is replaced by his child (because capital markets are imperfect, it does not make sense to sell the firm unless the child is especially untalented). With this they are able to explain a surprisingly large (up to 50%) fraction of the TFP gap across countries. However agents in their model have one period lives and follow a fixed bequest rule, so the possibility of undoing misallocation by accumulating resources does not arise. Adding the possibility of accumulating during your own life-time should weaken their effects. On the other hand there is no reason to take the idea that z is ability too literally: When a firm loses a contract because its contact in the buying firm has moved on, this is also a z shock, as is the introduction of a new product into the market. Interpreted in this way, the hypothesis of frequent and large z shocks does not seem prima facie implausible.

At this point the view of under-development based on misallocation has done enough to have earned a seat at the table. The next step is to flesh out this view and resolve the internal tensions that this paper has highlighted.

Appendix

Proof of Lemma 1

The proof follows the same steps as Theorem 1 in Dechert and Nishimura (1983). By way of contradiction, let $a > \hat{a}$ but $s_t(a, z) \leq s_t(\hat{a}, z)$. By utility maximization

$$v_t(a, z) = u[\pi_t(a, z) - s_t(a, z)] + \beta v_{t+1}[s_t(a, z), z] \geq u[\pi_t(a, z) - s_t(\hat{a}, z)] + \beta v_{t+1}[s_t(\hat{a}, z), z].$$

Likewise

$$v_t(\hat{a}, z) = u[\pi_t(\hat{a}, z) - s_t(\hat{a}, z)] + \beta v_{t+1}[s_t(\hat{a}, z), z] \geq u[\pi_t(\hat{a}, z) - s_t(a, z)] + \beta v_{t+1}[s_t(a, z), z].$$

will affect the extent of misallocation at the extensive margin and might amplify the output effects of the shocks.

Differencing the above two inequalities, we have

$$u[\pi_t(a,z) - s_t(a,z)] - u[\pi_t(\hat{a},z) - s_t(a,z)] \geq u[\pi_t(a,z) - s_t(\hat{a},z)] - u[\pi_t(\hat{a},z) - s_t(\hat{a},z)]. \quad (12)$$

Note that

$$[\pi_t(a,z) - s_t(a,z)] - [\pi_t(\hat{a},z) - s_t(a,z)] = [\pi_t(a,z) - s_t(\hat{a},z)] - [\pi_t(\hat{a},z) - s_t(\hat{a},z)].$$

But then the inequality in (12) contradicts the strict concavity of u (strictly decreasing differences).□

Proof of Proposition 1

Fix a z. Consider an agent with with initial wealth a_0. Denote his wealth and consumption by $\{a_t\}$ and $\{c_t\}$. Consider another agent with the same ability, but initial wealth $\hat{a}_0 < a_0$. Denote his wealth and consumption by $\{\hat{a}_t\}$ and $\{\hat{c}_t\}$. Denote the ratio of their marginal utilities by

$$\alpha_t \equiv \frac{u'(c_t)}{u'(\hat{c}_t)}.$$

Combining the Euler equations (9), we have that

$$\alpha_t = \alpha_{t+1} \frac{1 + r_{t+1} + \lambda(r_{t+1})\psi_{t+1}}{1 + r_{t+1} + \lambda(r_{t+1})\hat{\psi}_{t+1}} \leq \alpha_{t+1}.$$

The inequality follows because by lemma 1 $a_t > \hat{a}_t$, all t so that either $\psi_t = \hat{\psi}_t = 0$ or $\psi_t < \hat{\psi}_t$. The sequence $\{\alpha_t\}_{t=0}^{\infty}$ is nondecreasing and therefore converges on the extended real line. There are only two possible cases.

CASE 1: $\{\alpha_t\}_{t=0}^{\infty}$ converges to some $\alpha^* < \infty$. This is only possible if the sequence of multipliers $\{\hat{\psi}_t\}_{t=0}^{\infty}$ converges to zero, implying the desired result.

CASE 2: $\{\alpha_t\}_{t=0}^{\infty} \to \infty$. This is only possible if $\hat{c}_t \to \infty$ or $c_t \to 0$. But this would imply that

$$\lim_{t\to\infty} v_t(a_t, z) < \lim_{t\to\infty} v_t(\hat{a}_t, z).$$

Since wealth trajectories do not cross and the value function $v_t(\cdot, z)$ is weakly increasing, this is a contradiction.□

Proof of Proposition 2

Consider a steady state with some interest rate r (not necessarily equal to ρ). Consider first the case where *every individual* is in steady state. Then from the Euler equation a steady state with positive wealth has to satisfy

$$1 + \rho = 1 + r + \lambda(r)\psi(a, z) \tag{13}$$

We next want to argue that (given that the interest rate is constant), this steady state is stable from the perspective of an individual. This can be done using the apparatus of Skiba (1978), Dechert and Nishimura (1983) and Buera (2008). We don't include the detailed argument due to space restrictions. Suffice it to that individual wealth sequences are monotonic because policy functions are strictly increasing (Lemma 1), and that individuals either converge to a positive steady state satisfying (13) or decumulate wealth until it reaches zero. By a similar argument as in the diminishing returns case, we can rule out the case $r \neq \rho$: with $r > \rho$ the aggregate capital stock would grow; with $r < \rho$ everyone would be constrained and there would be no lenders. Again as above, all multipliers $\psi(a, z)$ must equal zero; otherwise constrained agents would be accumulating wealth. Therefore there is no misallocation on the intensive margin.□

References

Aghion, P. and P. Bolton (1997). A Theory of Trickle-Down Growth and Development. *The Review of Economic Studies 64*, 151–172.

Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics 109*(3), 659–684.

Alfaro, L., A. Charlton, and F. Kanczuk (2008). Firm-Size Distribution and Cross-Country Income Differences. *NBER Working paper 14060*.

Angeletos, G.-M. (2007). Uninsured Idiosyncratic Investment Risk and Aggregate Saving. *Review of Economic Dynamics 10*, 1–30.

Angeletos, G.-M. and L.-E. Calvet (2006). Idiosyncratic Production Risk, Growth and the Business Cycle. *Journal of Monetary Economics 53*, 10951115.

Bai, C.-E., C.-T. Hsieh, and Y. Qian (2006). The Return to Capital in China. *Brookings Papers on Economic Activity* (2), 61–88.

Banerjee, A. (2003). Contracting Constraints, Credit Markets and Economic Development. In L. Hansen, M. Dewatripont, and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Cambridge University Press.

Banerjee, A. and E. Duflo (2005). Growth Theory through the Lens of Development Economics. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Development Economics*, pp. 473–552. Elsevier B.V.

Banerjee, A. and E. Duflo (2008). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *mimeo, MIT*.

Banerjee, A. and A. F. Newman (1993). Occupational Choice and the Process of Development. *The Journal of Political Economy 101*(2), 274–298.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2008). Cross Country Differences in Productivity: The Role of Allocative Efficiency. *mimeo, University of Maryland*.

Buera, F. (2008). Persistency of Poverty, Financial Frictions, and Entrepreneurship. *mimeo, Northwestern University*.

Buera, F., J. Kaboski, and Y. Shin (2008). Finance and Development: A Tale of Two Sectors. *mimeo, Northwestern University*.

Buera, F. J. and B. Moll (2009). A Simple Model of the Wealth Distribution and the Macroeconomy. *Work in progress*.

Caselli, F. and J. Feyrer (2007). The Marginal Product of Capital. *Quarterly Journal of Economics 122*(2), 535–568.

Caselli, F. and N. Gennaioli (2005). Dynastic Management. *mimeo, London School of Economics*.

de Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to Capital: Results from a Randomized Experiment. *Quarterly Journal of Economics 123*(4), 1329–72.

Dechert, W. D. and K. Nishimura (1983). A Complete Characterization of Optimal Growth Paths in an Aggregated Model with a Non-Concave Production Function. *Journal of Economic Theory 31*, 332–354.

Galor, O. and J. Zeira (1993). Income Distribution and Macroeconomics. *The Review of Economic Studies 60*(1), 35–52.

Hsieh, C.-T. and P. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *forthcoming, Quarterly Journal of Economics*.

Jeong, H. and R. Townsend (2007). Sources of TFP Growth: Occupational Choice and Financial Deepening. *Economic Theory 32*, 179–221.

King, R. and S. Rebelo (1993). Transitional Dynamics and Economic Growth in the Neoclassical Model. *The American Economic Review 83*(4), 908–931.

Moll, B. (2009). Creditor Rights, Inequality and Development in a Neoclassical Growth Model. *mimeo, University of Chicago*.

Piketty, T. (1997). The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing. *The Review of Economic Studies 64*, 173–189.

Restuccia, D. and R. Rogerson (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics 11*(4), 707–720.

Skiba, A. K. (1978). Optimal Growth with a Convex-concave Production Function. *Econometrica 46*(3), 527–539.

Swan, A. (2008). New Evidence on the Marginal Product of Capital. *mimeo, Australian National University*.

Udry, C. and S. Anagol (2006). The Return to Capital in Ghana. *American Economic Review Papers and Proceedings 96*(2), 388–393.