

Brookings Global Economy and Development Conference
“What Works in Development? Thinking Big and Thinking Small”

**Big answers for big questions: the presumption of
macroeconomics**

**Abhijit Vinayak Banerjee
MIT
May 21, 2008**

Don't we know that all that matters for reducing poverty is growth, especially after China? And therefore we development economists should focus on the things that make growth happen: Macro policy and creating the right institutional environment. And not bother with the micro evidence...

No, no, and, as the expression goes, no. Every step in that syllogism is wrong, and, I will argue in this essay, each step is probably more obviously wrong than the previous one. But let us come to it step by step.

I. Growth is all you need

The one claim among the three that is both the most controversial and the least obviously false is the first one (this says something about the nature of macroeconomic debates): This is not because it is actually true---it is just that it starts from a claim that at least has a chance of being right.

What the cross-country evidence does seem to show (though this is disputed and I am no expert on the nuances of cross-country data) is that there is *no evidence* for a trade-off between growth and poverty reduction. Poverty goes down by more in countries that grow faster, and indeed even the elasticity of poverty reduction with respect to growth does not go down significantly in countries that grow faster.

This is certainly a useful correlation to know, especially given the amount of sententious rubbish that gets written justifying bad macroeconomic policies in the name of poverty reduction. But by itself it tells us very little about what we should or shouldn't do: One could, for example, read it to say that countries that do poverty reduction more effectively grow faster. Isn't that much less plausible than the opposite, you might say: After all, it is hard to imagine fast growth not reducing poverty. But how about a mixture of the two: Policies that cause poverty reduction makes growth happen, which in turn brings about more poverty reduction. After all, as Benabou and others have shown, countries that

redistribute more grow faster (though of course, the causality could be running from growth to redistribution).

Or how about a third view: that growth that is not accompanied by substantial poverty reduction is unsustainable (for political or other reasons) and therefore the episodes of sustained growth that we observe in the data are actually exactly the episodes where growth came with poverty reduction. In the relevant sense therefore, it is poverty reduction that determines growth and not the other way around.

Moreover, as Martin Ravallion has emphasized, the fact that on average countries that grow faster reduce poverty more, is in some ways less interesting than the fact that the effect of growth on poverty is so much greater in some countries than others. The point is that there is no reason to assume that this entire difference is just random: Indeed it is well recognized that the impact on poverty reflects the way the income was initially distributed---growth moves more people across the poverty line if there are more people near it to start with. But there are also some more interesting reasons why countries differ that have to do with policies and/or the initial distribution of skills and other endowments (which may be the result of policy as well). Indeed even in China, which is everyone's favorite example of poverty reduction through growth, a very substantial part of the massive poverty reduction that happened in the 1990s, was a result of one-off changes in the taxation of agriculture and the pricing of agricultural products (Chen and Ravallion).

II. Hence development economists should study growth

The problem with this comes down to basic economics: The one thing that everyone learns in their first economics class is that it all comes down to where your marginal product is highest. Even if growth were the best way to reduce poverty, we economists might want to focus on poverty reduction through other means if we think that is where we have the highest marginal product.

The reason why this is more than just a debating point, is that we know precious little about how to make growth happen. Even in terms of just accounting for growth, the remarkably optimistic results of Mankiw, Romer and Weil (1992) claiming that differences in savings rates, rates of investment in human capital and population growth rates can explain nearly 80% of the variation in GDP levels, has now been replaced by a more pessimistic position, summarized in Caselli (2005), which suggests that nearly two-thirds of the variation remains unexplained. The difference comes mainly from replacing the Mankiw, Romer and Weil measure of investment in human capital (a fraction of the population who have some secondary schooling) with a more continuous measure that gives weight to primary and tertiary schooling as well (Klenow and Rodriguez-Clare).

However even if we were able to explain away all the cross-country differences, it would tell us relatively little about how to narrow the existing differences. Savings rates, human capital investment rates and population growth rates are at least as much symptoms of the problem as their cause, and the challenge is in part to figure out how to move them in the right direction. For this we need to be able to identify proper causal factors.

The challenges of identifying causal factors using cross-country data are well-known: There is the fact that almost everything at the country level is a product of something else---educational investment, for example, to take one factor that was emphasized in the early cross-country literature (Barro), is clearly in part a product of the effectiveness of the government as a provider, which presumably has other effects as well, and in part a result of people's expectations about growth itself (Bils and Klenow). Moreover both countries and country policies differ in so many different ways, that in effect, every cross-country regression runs the risk of working with negative degrees of freedom. That is to say, there are so many different ways one could run the regression, that one could more or less guarantee that some of those regressions show significant "causal" effects, even in a world where policy is entirely irrelevant.

That does not mean that we have not learnt anything from this literature. In particular some of the more remarkable findings in the literature are convincing because, *a priori*,

they seem so entirely implausible. The most striking of these is almost surely the finding in Acemoglu, Johnson and Robinson (2001, 2002) showing that countries where in the initial years of European colonization settler mortality was high tended to be places which are still doing badly. Moreover it is not because these places were somehow handicapped from the start: There was actually what they call a “Reversal of Fortune” in this period---the countries that were the most prosperous at the time of colonization (at least measured by population density) actually end up at the bottom of the heap.

Acemoglu, Johnson and Robinson argue that this reflects the power of institutions: The places that were empty to start with and where settler mortality was low (which tended to be often the same places, for fairly conventional epidemiological reasons), were the places where the Europeans settled in large numbers and got the institutions that the Europeans were then developing, that would eventually provide the basis of modern capitalism.

It is certainly true that settler mortality is an excellent predictor of the quality of contemporary institutions at the country level, measured by, say, the risk of your capital being expropriated in that country. It is also true that these institutions as predicted by settler mortality are extraordinarily powerful in predicting the economic status of the country today. While this does not prove that it is these institutions that are responsible for growth (could be culture or political traditions, for example), it does say that some very long-run factors have a lot to do with economic success.

What does all this tell us about policy? We learn that institutions matter (perhaps) but not whether it would help to set up a particular set of institutions today: The evidence emphasizes the part of the difference in institutional quality that is attributable to things that happened several hundreds of years. Does that mean that institutions need to be developed over several hundred years for them to be effective (after all, the US constitution of today, strict constructionists notwithstanding, is a very different document today than it was written, enriched by two hundred years of jurisprudence, public debate and popular involvement)?

What is more worrying is that once we control for the cross-country variation in institutional quality predicted by differences in settler mortality, the standard measures of macroeconomic policy (monetary policy, trade policy, government involvement in the economy) have little or no effect on how fast countries grow (Dani Rodrik, Arvind Subramanian and Francesco Trebbi (2002)), implying not quite that policy does not matter, but at least that there is no evidence that the variations in policy that we see between countries of similar institutional quality make any difference to growth. In other words conventional notions of macro policy might matter, but perhaps only when the relevant variation in policy is the kind that we see when we compare the Singapores of the world with the Sudans: It is only the hyperinflations, the large-scale nationalizations and the civil wars that are obviously contra-indicated.

This does not say anything about unconventional forms of macro policy: It will probably surprise no one that no two episodes of sustained fast growth have been exactly alike and each successful country has pursued its own idiosyncratic policies as a part of its development strategy. In each of these cases the debate has been whether growth happened because of the particular deviant policy (industrial policy in Korea and Japan, massive government involvement in factor markets in China, as well as an undervalued exchange rate, forced savings in Singapore) or in spite of it. And in each case, predictably, the discussion has been inconclusive, though everyone seem to agree that there are other things these countries could have done that are worse.

Does that mean that they just lucked out: Or is there actually a lesson to be learnt from their success? The problem is that we do not know of any way to begin to answer that question based on the standard empirical methodologies used in the growth literature--let alone actually have something reliable to say about it--since these are essentially unique events and it is not clear what we could compare them to. We will return to this methodological challenge in the penultimate section.

By contrast, we know a significant amount about many specific strategies that can help improve the lot of the poor, and perhaps more importantly, we know how to go about learning more. One great advantage of randomized trials is that we can often start from a specific policy question and then look for the evidence: While it is not always feasible to get an implementing organization to implement the exact experiment that you want, there is so much going on these days, especially in the world of NGOs, that there is usually an opportunity to do something closely related. The fact that experimental work has now acquired a certain currency, helps very much here--there is much less resistance to the idea now outside academia than there was ten years ago, which means that there are many more opportunities.

Moreover there is little dispute that a carefully designed experiment gives us an unbiased estimate of the impact of the particular intervention being studied. To use the accepted jargon, experimental results are internally valid. This is in contrast to most forms of empirical research, where the internal validity is never conclusive; the one exception being certain natural quasi-experiments, where it is clear that some accident of fortune generated something very close to a real experiment.

The concern is with external validity: Many experiments are based on variation that is essentially local. In part this reflects the scope of the implementing organization, in part the fact that it is managerially easier (and therefore cheaper) to collect high quality data in a relatively small area.

The concern that this creates is whether the results from the experiment can be applied to settings other than where it was generated. There are really two parts to this objection. The first is environmental dependence---the results may be specific to the setting where the experiment was carried out. This is of course an old concern with all forms of empirical reasoning as David Hume, the Scottish economist and philosopher, pointed out more than two centuries ago. But it may be more serious here than in many other instances because we know that the setting matters and we only observe the result from one very particular setting. One potential advantage of cross-country research and macro

empirical research more generally (comparing regions within the same country, etc.), is that the “treatment” effect is an average across a large number of settings and therefore more generalizable.

It is tempting to go from this observation to posit a certain symmetry: All forms of learning have their limitations (true enough). Hence there is no reason to privilege one kind of evidence over another---cross-country regressions may teach us more or less than an experimental estimate—and therefore the evidence on growth policy is no worse than the evidence on specific strategies to fight poverty (misleading). To see where the symmetry breaks down, consider a scenario where we are worrying about generalizing from a certain experimental result. Say, we suspect that there are locations where a particular treatment may seem *a priori* appropriate but does not actually work. The point is that this hypothesis is entirely testable: All we need to do is to carry out additional experiments in different locations—indeed continue to experiment till we are either satisfied that the treatment effect is reasonably stable or find that it is not and therefore need to rethink the generalizability of the treatment. If we have a theory that tells us where it might break down, we focus the extra experiments there. If not, we choose random locations within the relevant domain.

This is more than an articulation of a principle. A number of experiments results have been replicated: For example, Bobonis, Miguel and Sharma get the same kind of impact of deworming (treatment for intestinal helminthes) on school attendance in North India that Miguel and Kremer found in Kenya, and Bleakley finds similar results using natural data from the US south in the early part of the 20th century using a natural experiment approach. Other results turn out not to be replicable: An information campaign that mobilized parent’s committees on issues around education and encouraged them to make use of a public program that allows school committees to hire local teachers where the schools are over-crowded, had a positive impact on learning outcomes in Kenya but not India.

The point is not that every experimental result generalizes. But there is always (at least in principle) a clear process that can inform us about the generalizability of a given result. Whether or not that process gets followed depends on a variety of practical concerns, such as the cost of doing another experiment and the cost of a potential delay, and it may not always happen, but if we cared enough about being right we could always do it.

Contrast this with the results from a cross-country regression. Suppose we suspect that the relation between women's literacy and fertility rates estimated from a cross-country regression is contaminated by the fact that women's literacy is correlated with many other unobserved aspects of woman's literacy. Running the same regression in different data sets many times over (regions within the same country, comparing the same region over time), may not help here, because every single one of these results may be biased by the presence of the same unobservable. The only way to solve this problem is to look for variation in women's education that is quasi-randomly or randomly assigned.

It is in this sense that quasi-experimental evidence clearly dominates. There is however, as already mentioned, a second, less obvious, problem with micro-experimental results: Consider what would happen if we try to scale up a program that shows, in a small-scale experimental implementation, that economically disadvantaged girls who get vouchers to go to private schools, end up with a better education and a higher incomes. When we scale up the program to the national level, two challenges arise: one is that there will be crowding in the private schools and the other is that the returns to education will fall because of increased supply. For both reasons the experimental evidence would overstate the returns to the vouchers program.

These so-called general equilibrium effects pose a problem that has no perfect solution. We could try to do experiments at a larger scale so that these kinds of spillovers are internalized within the experimental units. But this is expensive and often infeasible. However there are clearly many instances where this is only a problem if we are not careful in interpreting the results of our experiments. For example, the results of an experiment that says that in the current state of the health system a particular intervention

can increase immunization rates by x percent, tells us what we would get if we universalized the intervention *under the assumption that nothing else changes in the immunization environment*. In other words, the government is not allowed to say that it wants to boost immunization rates without spending an extra penny, or rather, if that it is what it wants, then the relevant interventions may have to be very different.

The real problem arises when keeping the environment constant is not directly an option. For example, while the government could presumably solve the problem of the supply of private schools that we brought up above by making it sufficiently lucrative to set up schools, the experiment tells us nothing about the price it would have to pay for that. The difference with the immunization case comes from the fact that the unit cost of expanding immunization is likely to be more or constant at least in the relevant range (immunization does not require that much extra man-power and it is easy to procure more vaccines on the world market, given enough lag time). Or it may be that it does not make economic sense to keep the environment constant: The fact that the price of skill goes down when the supply goes up, may be exactly what makes the program socially desirable.

The response to this however is not less micro evidence but more. The problem arises because we do not know the relevant elasticities of supply and demand, and more generally, the production functions. Had we known them, we would have mapped the increase in demand for private schooling and the rise in the supply of skills through them, to infer the true social benefits of the intervention. Simply implementing the intervention without doing that obviously runs the risk of a disaster, but the alternative of not doing anything because we do not know these elasticities seems unnecessarily pessimistic. For the long haul we clearly need more carefully designed experimental/quasi-experimental studies that will give us the relevant elasticities.

The point once again is not that the experimental/quasi-experimental route always offers us an answer, but that it offers a process by which we should be able to converge to useable evidence. Of course there is a long way to go, but the process relatively well-defined, and we know we are headed. And while the ambition is limited, it may be

achievable: We might soon know enough in certain domains (health, education, environment) to be fairly confident that lives are being saved and people are living better lives because of the evidence we have.

III. Studying growth means studying macro evidence

This is the illusion of commensurability: Big questions must have big answers. Growth is surely the biggest question that we economists tackle. Hence the evidence that can inform growth policy must be evidence about big things

There are at least two senses in which this is misleading: First suppose the conclusion from the macro evidence is that reducing corruption is vital for promoting growth. But reducing corruption how? And what form of corruption are the most worth fighting?

There is a very simple reason why we cannot use macro evidence to answer these kinds of questions: There is no comprehensive data set on corruption, because even though there are many data sets that record the impression of corruption (usually among businessmen) associated with particular countries, this evidence does not sort between the various sources of unhappiness among the victims of corruption (coercion, unnecessary delays, meaningless procedures¹) but even more importantly, it says little about the many costs of corruption that are not observed by the individual business-man (misallocated resources, biased decision-making, lost government revenues, lobbying and other forms of rent-seeking, investments that did not happen). Therefore we need micro data to tell us which of these forms of corruption are the most costly.

The situation with respect to how to fight corruption is even worse: There is no macro data on fighting corruption, for the simple reason that different countries have chosen their own distinct ways to deal with corruption. To make matters even more complicated, the evidence from the experimental evidence suggests that the effectiveness of anti-corruption strategies turn very much on the details: Research by Ben Olken (2007) in

¹ This is where the World Bank's "Doing Business" Reports have been a useful innovation.

Indonesia finds that a strategy of inviting villagers to a meeting where village road construction expenses were discussed reduced “missing” expenditures by a third or so when the invitations to the meeting were distributed to school children (to take home to their parents) but not at all when the village head got to distribute them. Olken, plausibly, interprets this as an effect of popular participation: The village heads managed to make sure that the meetings were stacked when they controlled the invitations, but not when it was taken out of their hands.

There is obviously no way one would pick up this fact from macro data, simply because this type of fine variation never shows up there. And while the conclusion seems plausible *ex post* (of course it is better to distribute them in school) how did we know that the head could not discourage/intimidate people enough to keep them out of the meeting even when they had been invited? After all, the meetings were always open to the public but no one showed up except when they were specifically invited.

The same point can be made, mutatis mutandis, about many other dimensions of growth policy: Investment in education (where? at what level? through better teacher training or greater parental involvement?), investment in health-care, more effective capital markets, etc. In the end, details matter too much for it to be possible to do effective growth policy without experimental/quasi-experimental data.

The second reason why macro data tells us little about how to do growth has already been discussed: Countries seem to vary enormously in their performance even after we control for differences in standard macro factors like capital, human capital, demographics, etc., and differences in macro measures of policy does not help much here either, at least once we control for differences in institutional quality.

How do we enter the black-box of TFP differences: Why does India in 1990s seem to have a TFP level that is about half of that in the US if we use Lucas’ Cobb-Douglas model of GDP (Banerjee and Duflo (2005)? One answer to this question that was proposed in the growth literature is that the TFP differences are the result of spillovers

from human capital investments. Human capital investments are worth much more than standard growth models (like the one we used) have given them credit for, because they also benefit everyone in the country. Therefore relatively small differences in human capital can explain big differences across countries.

The problem with this theory is that it fails the quasi-experimental test. Both Acemoglu and Angrist (XXXX) and Duflo (XXXX), who look at this question using micro data from credible natural experiments in the US and in Indonesia respectively, which lead to expansions in high school attendance and completion rates, find that if there are spillovers they are swamped by the standard diminishing returns effects.

Another standard theory of growth attributes the TFP difference to differences in access to effective and appropriate technologies. Yet as Banerjee and Duflo (2005) argue, once again the micro data gets in the way. They argue that if the entire two-fold difference in TFP between India and the US has to be fully explained by technological differences then, given the 1-1.5% growth rate of TFP in the US, India in 2000 should have been using technologies developed in the 1950s. In fact, the best Indian firms use technologies that are entirely contemporary and a recent Mckinsey report concludes that upgrading very close to the latest technologies would be profitable at the current factor prices in India. Hence it is not access to viable technological improvements that seem to be the constraint on the average Indian firm.

At least a part of the answer to the TFP puzzle seems to come from massive misallocation of resources within the same economy, something that is not picked up by any of the macro aggregates that are used in growth accounting exercises. These misallocations are not the product of any one distortion but rather the cumulative effect of many, many individual distortions resulting from both government failures and market failures. Banerjee and Duflo (2005) describe the evidence for these distortions in some detail drawing on range of micro-studies. They then carry out a heuristic exercise to assess whether the extent of observed misallocation is large enough to explain away the Indo-

US TFP differences. Their answer, which they propose quite tentatively, since what they do is no more than a finger exercise, is yes: If we are willing to assume a model where there is some increasing returns at the firm level, the fact that the medium firms in India are too small and too numerous relative to what they would be in an efficient economy, can actually explain the entire TFP gap.

Hsieh and Klenow (2006) use data from firm-level annual surveys from the US, China and India to carry out a much more empirically founded version of the same exercise. They calibrate a model of monopolistically competitive differentiated firms using this data and show that the allocation of resources across firms within the same industry is indeed much more distorted in both India and China than in the US, and that in particular it is most productive firms that are too small in both those countries. If these countries could achieve US-level efficiency in the allocation of resources within the same industry, they calculate, TFP would go up by 30-45% in China and 45-50% in India. Clearly there may also be misallocation across industries, which would presumably add to this total.

This is all based on micro evidence but not necessarily data from any experiment/quasi-experiment (though there are those too): Indeed a lot of the data is simply descriptive and does not require a causal interpretation. However anyone looking at the economy through a macro lens would miss them.

This research is too new to have many specific implications for policy. But it suggests a very different view of what we are looking for: The range of distortions are so diverse that it is hard to imagine that we would not want to address them separately and they seem specific enough that we can at least think of addressing them through policy. For example, it seems clear that fertilizer is massively underused in parts of Africa: the question is to what extent this is a result of an unwillingness to take risks, the unavailability of credit, the lack of the right internal or external incentives for long-range planning, distortions in the land market or a lack of understanding of the benefits of fertilizer. This is the kind of problem that is probably best addressed by a combination of theoretical thinking and experimental work, exemplified by the work of Duflo, Kremer

and Robinson (2007) on fertilizer adoption in Kenya. Their results clearly suggest that a part of the problem is a lack of ability to commit to a long-range plan and that a simple contract that solves this commitment problem can massively increase fertilizer adoption.

This is no doubt only a part of the answer: And it is possible (though hardly obvious—there many NGOs in Kenya) that there is not enough implementation capacity in Kenya to provide the needed contract to everyone. On the other hand, once the need is well-understood, there is no obvious reason why the market would not start to offer it. This is a possibility that Duflo, Robinson and Kremer are now investigating.

Beyond this, there is the concern that in a relatively fragile state like Kenya, there is no point in trying to do anything about agricultural productivity because all gains will eventually be destroyed by some form of civil conflict. This is both much too pessimistic--Kenya has been growing quite fast for the last few years now---and potentially self-fulfilling, since it may be precisely the lack of any economic progress that will eventually lead to civil conflict.

Which brings us to our last, most radical, thought: It is not clear to us that the best way to get growth is to do growth policy of any form. Perhaps making growth happen is ultimately beyond our control. Maybe all that happens is that something goes right for once (privatized agriculture raises incomes in rural China) and then that sparks growth somewhere else in economy, and so on. Perhaps, we will never learn where it will start or what will make it continue. The best we can do in that world is to hold the fort till that initial spark arrives: make sure that there is not too much human misery, maintain the social equilibrium, try to make sure that there is enough human capital around to take advantage of the spark when it arrives. Social policy may be the best thing that we can do for growth to happen and micro-evidence on how to do it well, may turn out to be the key to growth success.