EPW Commentary October 1, 2005

Is There Too Little Theory in Development Economics Today?

Dilip Mookherjee

Development economics is increasingly becoming an empirical discipline today. Where theory reigned during the 1980s and early 1990s, in the last 10 years or so the primary research concerns have become increasingly driven by empirical and policy issues. In this essay I wish to reflect on the balance between theory and the rest in the discipline, thinking aloud about the question of whether the pendulum has shifted away a bit too far, and whether there is need for some reorientation of the dominant research agenda or methodology.

In the next section I will revisit classic "theory versus empirics" debates involving some of the great 20th century economists, and distil from this a notion of alternative stages in the maturation of any given field of economics. Following this I will attempt to place the evolution of methodology in development economics within this general scheme, in order to gain some perspective of where the field currently stands, and the role of theory at the current frontiers of research.

Given the nature of this symposium, I will attempt to be brief and succinct, rather than embark on an expansive and scholarly essay. Accordingly I will eschew accuracy and detail for the sake of breadth and brevity. Indeed, I see this as an opportunity to air and thus clarify my own subjective views, and invite feedback from others.

Classical "Theory versus Empirics" Debates

Regarding the appropriate balance between theory and empirics, the great economic thinkers of the first half of the 20th century had strikingly diverse views. In his preface to his monumental History of Economic Analysis, a thousand page historical account on the evolution of economic theory, Joseph Schumpeter expressed the judgment that the study of economic history and economic statistics should precede theory. At the other extreme John Maynard Keynes maintained a profoundly distrustful attitude towards statistics. Towards the beginning of his well known review of Jan Tinbergen's monograph on statistical testing of business cycle theories, Keynes (1939) writes:

Professor Tinbergen is obviously anxious not to claim too much. If only he is allowed to carry on, he is quite ready and happy at the end of it to go a long way towards admitting, with an engaging modesty, that the results probably have no value. The worst of him is that he is much more interested in getting on with the job than in spending time in deciding whether the job was worth getting on with. He so clearly prefers the mazes of arithmetic to the mazes of logic, that I must ask him to forgive the criticisms of one whose tastes in statistical theory have been, beginning many years ago, the other way around (p 559).

He then goes on to list various possible sources of bias, which in the vocabulary of modern econometrics would translate into problems of omitted variables, errors in variables, multicollinearity, simultaneity, incorrect functional form, time lags, trend corrections, parameter instability, and problems in extrapolation. Gathering these together, he expresses his scepticism concerning the usefulness of regression analysis in the process of inference concerning the validity of theories:

Thirty years ago I used to be occupied in examining the slippery problem of passing from statistical description to inductive generalisation in the case of simple correlation; and today in the era of multiple correlation I do not find that in this respect practice is much improved (p 566).

The review ends thus:

I hope that I have not done injustice to a brave pioneer effort…I have a feeling that Professor Tinbergen may agree with some of my comment, but that his reaction will be to engage another ten computers and drown his sorrows in arithmetic. It is a strange reflection that this book looks likely, as far as 1939 is concerned, to be the principal activity and raison d'etre of the League of Nations (p 568).

Fortunately Tinbergen did not give up, and neither did the fledgling field of econometrics, despite the rapid demise of the League of Nations. Tinbergen's patient and detailed reply to Keynes reads like the rejoinder that any modern econometrician would probably have written [Tinbergen 1940]. Specifically, that many influences could be excluded from the estimated regression and collected as a residual, suitable assumptions regarding which would validate the inference, assumptions that can themselves be tested from the data; that multicollinearity is only a problem of degree not of kind; that non-linear functional forms can be approximated by polynomials, and so on. The rejoinder hardly affected Keynes' scepticism, whose subsequent reply ended by labelling econometrics as "statistical alchemy" rather than a science [Keynes 1940].

Yet it is undeniable that econometric analysis has progressed impressively since the time of the Keynes-Tinbergen debate, particularly with respect to combating the very problems that they discussed. The scientific status of econometrics is far less in doubt today, an issue argued persuasively by David Hendry (1980) in his inaugural lecture at the London School of Economics in 1980:

Forty years after Keynes wrote, his review should still be compulsory reading for all who seek to apply statistical methods to economic observations...To Keynes' list of problems, I would add stochastic mis-specification, incorrect exogeneity assumptions…, inadequate sample sizes, aggregation, lack of structural identification and an inability to refer back uniquely from observed empirical results to any given initial theory.
That the subject is exceedingly complicated does not entail that it is hopeless. Considerable progress has been made on the technical aspects, such as studying the consequences of the various problems just listed, designing means of detecting these, developing methods that mitigate some of their ill effects or handle several complications at once, and analysing the properties of estimators when the sample size is small…Much of this technical work is essential background to understanding and correctly interpreting empirical findings and, although some work may have turned out to be otiose in retrospect, the ever increasing level of technique is not a symptom of alchemy (p 396).

I doubt that many economists today would disagree with this assessment. The contemporary relevance of the Keynes-Tinbergen debate to development economics is confined to those first generation empirical and field studies that are essentially descriptive accounts, yet aspire to have relevance for inferring the validity of some (frequently implicit) theory of the development process, predicting the future, or recommending suitable policies. Many such studies aim to avoid the latter

explicitly, in the standard anthropological approach of “description for description’s sake” – yet often make implicit inferences and policy judgments from their descriptions. Description is one thing, but inference is quite another – what Keynes referred to as the “slippery problem of passing from statistical description to inductive generalisation”. This is exactly the purpose of modern econometrics, i e, to explicitly specify the hypothesis to be tested, derive its observable implications, use these in inference, and test as far as possible assumptions required for the validity of the inference. Such studies are now common at the frontier of contemporary empirical research in development economics.

A decade later a similar debate arose between Tjalling Koopmans and Rutledge Vining, following Koopmans’ critical review of an empirical analysis of business cycles by Burns and Mitchell.[1] Koopmans’ critique provides an eloquent statement of the shortcomings of “measurement without theory”, drawing analogies from the relationship between the empirical Keplerian stage and theoretical Newtonian stage in celestial mechanics. He provides three main arguments.

My first argument, then, is that even for the purpose of systematic and large-scale observation of such a many-sided phenomenon, theoretical preconceptions about its nature cannot be dispensed with…(their) analysis employs the following seven series… There is no systematic discussion of the reasons for selecting these particular variables as most worthy of study...The lack of guidance from theoretical considerations is perceivable also in the choice of the measures computed from the variables selected…The rejection of the help that economic theorising might give leaves a void. For now there is a need for some organising principle to determine on what aspects of the observed variables attention should be concentrated… (pp 190-91).
The movements of economic variables are studied as if they were the eruptions of a mysterious volcano whose boiling cauldron can never be penetrated. There is no explicit discussion at all of the problem of prediction, its possibilities and limitations, with or without structural change, although surely the history of the volcano is important primarily as a key to its future activities. There is no discussion whatever as to what bearing the methods used, and the provisional results reached, may have on questions on economic policy. This, then, is my second argument against the empiricist position: without resort to theory, in the sense indicated, conclusions relevant to the guidance of economic policies cannot be drawn (pp 196-97)
...the extraction of more information from the data requires that, in addition to the hypotheses subject to test, certain basic economic hypotheses are formulated as distributional assumptions, which often are not themselves subject to statistical testing from the same data. Of course, the validity of information so obtained is logically conditional on the validity of the statistically unverifiable aspects of these basic hypotheses. The greater wealth, definiteness, rigour, and relevance to specific questions of such conditional information, as compared with any information extractable without hypotheses of the kind indicated, provides the third argument against the purely empirical approach (p 200).

Vining’s response is rich and detailed, which among various other points refers back to a classification of different stages of research that originated with Trygve Haavelmo:

Haavelmo gives a rough fourfold classification of the main problems encountered in quantitative research: first, the construction of tentative theoretical models; second, the testing of theories; third, the problem of estimation; and fourth, the problem of prediction. It may be noticed that the first problem (possible the second, depending upon one’s interpretation) is the only one of the four that is not a problem of strictly modern statistical theory. He goes on to say that the “explanation’’ of phenomena “consists of digging down to more fundamental relations ... those that have a great degree of invariance or autonomy with respect to the ordinary or reasonably expected changes in economic structure, and a theory is a construction of a system of autonomous relations (p 211).

Vining then goes on to discuss the shortcomings of the prevailing theoretical paradigm and therefore argues that business cycle research was still in the first stage:

Is the Walrasian conception not in fact a pretty skinny fellow of untested capacity upon which to load the burden of a general theory accounting for events in space and time which take place within the spatial boundary of an economic system? When we think of the enormous body of factual knowledge digested and systematised in the other fields of variation and the meagreness of our own results from efforts to systematise, are we quite ready to leave Haavelmo's first problem and launch into the last three problems in estimation theory?...Burns and Mitchell presumably are still on Problem 1. Koopmans has vaulted over, some would say hastily if the research is in the field of economics, to Problems 3 and 4. This is not to deny the very great interest that economic research has in the results of Kopmans' group on Problems 3 and 4. But it is to express belief that, so long as a field of knowledge continues to develop, workers will be puttering around, and not in vain, within the unexplored expanses of Problem 1 – the searching for regularities and interrelations of regularities and the feeling around for interesting theoretical models. Not all this work will find formal mathematics of immediate use, and much of it will be such an explorative character as to render almost meaningless the notion of a planned maximisation of information from given data (pp 212-13).

Vining's rejoinder is an important reminder of the value of exploratory data analysis as a preliminary stage of research on any given economic phenomenon, which precedes and aids the formulation of concrete theories of that phenomenon. The large masses of data on planetary motion collected by Tycho Brahe were no doubt a useful step for Kepler in his effort to organise the observed phenomena into a number of empirical laws, which in turn were a precondition for the Newtonian stage. In most areas of economics where the relevant theory is far from obvious, the value of similar exploratory data analysis in helping the evolution of such theories cannot be understressed.

Accordingly I would modify Haavelmo's classification to add a pre-theory stage prior to his stage 1, and for purposes of brevity merge the second and third stages of testing and estimation. The modified classification would then constitute the following: Stage 1: empirical description of the relevant phenomenon, consisting of exploratory data analysis aimed at helping identify empirical regularities that need to be explained by a suitable theory, and in addition the nature of assumptions that such a theory can make without gross violation to the empirical patterns. Stage 2: the formulation of a relevant theory, including derivation of potentially observable (hence falsifiable) implications. Stage 3: The testing and estimation of theories, a stage which may lead back to modification or replacement of the previous theories, in an iterative back and forth with Stage 2. Stage 4: use of the least unsuccessful theory from the standpoint of empirical verification for purposes of prediction and policy evaluation. From the standpoint of this classification, we can interpret Koopmans' critique of empirical work as pertaining to Stages 3 and 4, while Vining's rejoinder can be interpreted as a defence of Burns and Mitchell's work as pertaining to the pre-theory Stage 1. In the next section I shall review the ebb and flow of the role of theory and empirical research in the field of development economics over the past century in the light of this particular classification.

The Role of Theory in Development Economics in the Past 75 Years

Barring important isolated contributions in the earlier part of the century such as the work of Allyn Young, development economics emerged as a distinct field in the 1940s and 1950s. The principal protagonists in the early decades combined both Stage 1 empiricists such as Kuznets and Myrdal, and Stage 2 theorists such as Rosenstein-Rodan, Hirschman, Leibenstein, Lewis, Nurkse, Scitovsky and Sen. The field remained quite aloof and isolated from mainstream neoclassical economics, as the principal ideas of that literature – coordination failures, poverty traps, take-off into self-sustained growth, critical minimum effort, pecuniary externalities, balanced and unbalanced growth, dualism, surplus labour, institutional failures – were not quite consistent with the Walrasian paradigm. Instead, they were based on implicit assumptions of pervasive externalities, missing markets, economies of scale, imperfect competition, and imperfect information, for which satisfactory theories comparable in logical completeness, detail and

elegance to the Walrasian theory were then lacking. The key policy concern was the suitable design of strategies of development planning led and regulated by the state. Later in the 1960s and 1970s the theories followed in a similar tradition, addressing issues such as cost-benefit analysis, migration and optimal growth, in the work of Lefeber, Little, Mirrlees, Chakravarty, Sen, Dasgupta, Marglin, Harris,Todaro and many others. During this stage theoretical research predominated, or gained more visibility, with limited interchange between theoretical and empirical researchers. The field was definitely at somewhere between Stages 1 and 2. Stage 3 involving detailed testing and estimation of the theoretical models had not yet begun.

The next phase involved a shift to a different set of phenomena and theoretical frameworks, which brought development economics closer to the mainstream. Key to this were theoretical developments in mainstream economics in the 1970s, which began to incorporate departures from the Walrasian paradigm systematically, especially with respect to theories of imperfect competition and of asymmetric information. On the one hand these led to rigorous theories of some of the ideas that the early pioneers had discussed in less formal terms, such as coordination failures, endogenous growth theory, incomplete markets and efficiency wages. The literature on asymmetric information and contract theory began to be applied to agrarian institutions in developing countries, motivated by the empirical patterns documented for the latter by Bardhan and Rudra in the 1970s. The 1980s saw a large range of theoretical contributions of this kind, a broadening of issues including dynamic implications of contractual imperfections, political economy, governance problems, and the role and functioning of social norms. It represented the maturation of Stage 2 research in the field.

Next Stage

This enabled Stage 3 to commence. As a backdrop to this it is worth mentioning changes to mainstream theory that had also occurred during the 1970s and 1980s. It could now accommodate a far broader range of phenomena, with the increasing incorporation of game theoretic tools. But this also meant that the Walrasian paradigm, a single overarching unified framework that had dominated economics for almost a century or more, was giving way to a fragmented collection of disparate models designed to explain disparate phenomena. Replacing price theory by non-cooperative game theory meant there was now scope for great arbitrariness in the formulation of the institutional setting: who are the players, what are their strategy sets, who moves when, knows what and so on. Added arbitrariness is introduced by a proliferation of equilibrium solution concepts in the presence of dynamics and imperfect information. More recently game theorists have been exploring yet other variations in terms of limited rationality and alternatives to common knowledge assumptions. The result is an embarrassment of riches when it comes to the choice of a theoretical model for almost any phenomenon.

This paradigm shift has of course occasioned many other methodological debates within the profession, most notable of which is the interchange between Carl Shapiro and Frank Fisher concerning models of industrial organisation.2 Their debate highlights the strength and weakness of the “new institutional economics” grounded in game theory. On the one hand one can now more easily formulate theoretical models of a much large ranger of phenomena; on the other, one can find a profusion of very distinct models for the same phenomenon, which are empirically difficult to distinguish. While Fisher laments that “there is no theory of oligopoly”, Shapiro hails the ability of the approach to capture the richness of forms of oligopolistic interaction and business strategy, analogous to theories of evolutionary biology grounded in general principles (of adaptation and survival) that yield a disparate range of behaviour and strategy in different environments.

Shapiro’s view is consistent with the now widely accepted view in development economics that “institutions matter”. But for the further progress of research, the profusion of theories for any given pheneomenon that are difficult to distinguish empirically, is a bewildering problem. What agents care about, what they can do, when do they move, do they interact repeatedly, are there

important state variables or is the interaction represented by a repeated game, how patient they are, what can they contract over – these are all choices which theoretical models today must confront, with very little guidance from empirical facts. So while the new theoretical developments provided a language to discuss many classic developmental issues, unlike the Walrasian paradigm, they raised at least as many new questions as they resolved. Attempts to extend the Walrasian paradigm to contexts of asymmetric information have similarly run into the problem of many competing formulations with diverging implications for inferences concerning market failure, and the role of the state. An instance of this is the absence of any single competitive theory of moral hazard, with Prescott and Townsend constructing one such theory where the classical Arrow Debreu theorems extend, while Stiglitz and his various co-authors construct alternative competitive models where they do not.3 These competing extensions vary with respect to the formulation of commodities that are traded, and of the agents in the economy, issues so fundamental that they cannot be empirically distinguished. Nor can they be settled on theoretical grounds either, leaving economists to trusting their own subjective judgment for the most appropriate formulation in any given setting.

Back now to Stage 3 of the progress of the research agenda in development economics, which can be said to be well under way over the past decade. The typical paper at any leading conference or journal in development economics is now of this genre. It will typically focus on an issue or phenomenon on which the authors have gathered some empirical data, frame a theoretical hypothesis which can be "taken to the data", followed by econometric estimation and testing. Even theoretical papers are pressed to refer more to documented empirical facts for purposes of motivation and justification. Increasingly scholars are going to the field to get a feel for ground level institutions, in order to decide on the most salient issues to focus on and the kind of theoretical formulation that is best suited to the task, then collect their own data, before proceeding to the econometric analysis. The theoretical model helps identify the relevant variables, ways to measure them, identify potential econometric problems such as endogeneity bias, measurement error, selection bias and so on, and then the best ways of testing for these biases. The empirical methodology is quite close to those pursued in other fields of economics, most notably labour economics. This is pretty much the kind of research that Koopmans had in mind while criticising the Burns and Mitchell statistical compendium of business cycles. To my mind this kind of empirical work represents an important stage of maturation of the field, an important progression from Stages 1 and 2 to Stage 3.

The key problems confronted by a typical empirical paper today concern the seriousness of the various possible econometric biases. The skill and quality of the paper is related to the degree of persuasiveness of the inferences made by the analysis, in terms of its vulnerability to inferential errors. Most of these problems can be confronted with appropriate econometric tools, but ultimately with the availability of higher quality data sets, e g, with better measured variables, longitudinal panels to control for unobserved heterogeneity, and availability of suitable "instruments" for potentially endogenous regressors and selection biases. These econometric concerns are important and substantive, and lie at the heart of Keynes' "slippery problem of passing from statistical description to inductive generalisation'', or going from observed correlations to inferences concerning causation.

Microscopic Research

Not surprisingly, all this comes at a price. Research papers tend to get evaluated almost exclusively in terms of their success in combating the econometric problems, often to the exclusion of the importance of the context or issues addressed by the analysis, the imaginativeness of the underlying hypotheses formulated or tested, or the importance of the findings from a wider standpoint. A well executed paper goes into a particular phenomenon in a particular location in considerable depth, data permitting. The research is consequently increasingly microscopic in character. We have very little sense of the value of what we have learned for any specific location to other locations. If suitable data does not exist to address a

given topic in an econometrically satisfactory manner, then the research project is dropped, even though some aspects of the data would be of wider interest.

The concern to minimise econometric bias is increasingly motivating the discipline to study randomised experiments, either natural or controlled. In such experiments the treatment of the key explanatory variable of interest is randomised by design, much like the statistical experiments initiated and applied to the context of agriculture crop experiments and more recently, clinical trials of drugs in medicine. This is a way of ensuring the observed treatment effects are entirely uncontaminated by any kind of bias, either owing to measurement error, exclusion of omitted variables or reverse causality. More often than not the exercise concerns the effects of some policy intervention on an outcome of intrinsic interest to policy-makers. In order to avoid the problem of vulnerability to any kind of arbitrariness with respect to the theoretical formulation of structural relationships, only "reduced form" relationships connecting the treatment variables to outcome variables are estimated. The result is akin to clinical trials concerning a particular medical treatment, the effect of which on some dimension of health outcomes of patients is sought to be estimated. Once the relevant "experiment" is set up – i e, the treatment and outcome variables identified, and the randomised selection of treatment and control groups implemented, all that remains to be done is a statistical comparison of corresponding averages of the outcome variables across the two groups. The purpose is not to understand the underlying structure of the system of relationships generating the outcomes, only the statistical outcome impact of certain policy treatments. Indeed, the only knowledge of structure required is that involved in designing the experiment. To this extent it responds to Koopmans' first critique, and by its very nature also to the second, being concerned directly with policy evaluation, though it can be argued (as I will below) that greater use of theory could permit a wider range of policy assessments.

Controlled experiments of course can be used for other purposes as well, such as in testing theories, as they are used frequently in the natural sciences. With few exceptions, the wave of recent randomised evaluations in development economics are not designed to test theories, but proceed directly to the question of policy evaluation that constitutes Stage 4 research. In particular, these represent an approach to evaluate policies with minimal knowledge or interest in structure. It is in this sense that there is little theory at the frontier of development economics today; this seems to be reflected in the curricula and training of graduate students at leading universities. Whether it is "too little" or "enough" is the question I want to finally address.

'Too Little' or 'Enough' Theory?

Let me now put forward the set of reasons why I feel that it is "too little". First, there are still many important and interesting theoretical questions still waiting to be addressed. These include dynamic implications of factor market imperfections for individual agents, for their investment decisions that affect the evolution of their families' future assets and occupations, and how these would be affected by external shocks or policy interventions. The interaction of economic processes with political institutions and social norms is another important area where many interesting theoretical possibilities remain to be explored. The implications of replacing traditional assumptions of rationality by behavioural models, and endogenous evolution of preferences for many key policy issues in development are just beginning to be explored.

Second, even within the context of randomised policy experiments, there is considerable room for enlarging the scope and precision of inferences if they were combined with an effort to test for alternative theories of the structure of underlying relationships generating observed outcomes. There are a variety of reasons for this. For one, observed outcomes do not tell us what it is that economic agents really care about, i e, the nature of their preferences. From a policy standpoint, eventually, welfare effects should be what really matter. How can we ascertain which kinds of interventions members of any given community really care about? Social scientists risk imposing their own judgments regarding this matter, given the fact that due to their expense and organisational difficulty they have to choose only a few interventions to evaluate. In the 1970s

and 1980s the main issues that development economists used to address in the context of rural areas of underdeveloped countries were imperfections in land and credit, issues that have since been superseded by the delivery of education and health services in contemporary research. There is relatively little attention to the comparative significance of investments in rural infrastructure, such as roads, irrigation and marketing arrangements, which may well be more fundamental. One accordingly needs a theory concerning the objective functions of agents, concerning their own evaluation of alternative public interventions, which needs to be tested.

Moreover, by their very nature (similar to many clinical drug trials) the scope of randomised experiments are limited to assessment of short-term effects, to ignoring possible side-effects or general equilibrium responses, and of a narrow range of policy interventions. Wider assessments would be possible in the presence of better knowledge of the underlying behavioural and structural relationships. For instance, theories about the pay-offs of agents, or the nature of technology and costs, will permit simulation of the effects of an alternative range of policies. One could enlarge the scope of randomised experiments to allow testing or calibration of structural models that would enable such simulations to be carried out.

Finally, we have now come to expect that institutions matter, and that institutions vary greatly across communities, regions and countries. If so, what we learn about the effects of a given policy intervention in a certain context may not apply to other contexts as well. The problem is obviously far more serious for economics than in clinical drug trials, since heterogeneity in human physiology is undoubtedly far less significant compared to economic and social institutions. Theories of institutions that relate them to observable community characteristics could help us understand how context specific the results of particular experiments are.

I should reiterate that my argument is not that randomised experiments do not represent a major step forward in the methodology of economists. Instead, it is that they tend to be conducted in a manner to minimise the use or testing of theoretical hypotheses. They could be substantially enriched if they were also used as a tool for testing theoretical hypotheses, or at least maximising the information that could be extracted from such experiments. In this respect they are still subject to Koopmans' third critique. I would add to this that understanding the structure of economic relationships is an objective of innate intrinsic importance in social science, quite apart from their usefulness in prediction and policy evaluation.

In summary, I am not quite sure that we have made sufficient progress in Stage 3 to proceed comfortably to Stage 4. Experimental evaluations represent an exciting new innovation in research methodology, but they could be designed to also test theories instead of just measuring the impact of specific policy interventions. Besides the prospect of greater interaction with empirical and experimental methods, investing in theoretical research is desirable for the sake of posing new questions and hypotheses that are bound to generate interesting spin-offs for future empirical and policy analyses. So my hope is that the field of development economics will pursue a more balanced portfolio of research methodologies in the future, with a little more emphasis on theory than is the current norm.

Email: dilipm@bu.edu

Notes

1 This debate originally appeared in the *Review of Economics and Statistics* between 1947 and 1949. It is reprinted in the American Economic Association's volume of *Readings in Business Cycles.*
2 This appeared in the *Rand Journal of Economics,* Vol 20, No 1, Spring 1989.
3 Elsewhere I have elaborated further on the relation between these two alternative formulations of competitive economies with asymmetric information, e g, 'Market Failure and Information' in Bhaskar Dutta (ed), *Welfare Economics,* Oxford in India Readings, 1993, Oxford University Press, New Delhi, and 'Markets versus State: A Sterile Controversy?' in Kaushik Basu, Pulin Nayak and Ranjan Ray (eds), *Markets versus Governments,* 2003, Oxford University Press, New Delhi.

References

Hendry, D (1980): 'Econometrics: Alchemy or Science?', Economica, Vol 47, November, pp 387-406.
Keynes, J M (1939): 'Professor Tinbergen's Method', *Economic Journal*, Vol 49, No 195, September, pp 558-77, Review of *A Method and Its Application to Investment Activity* by Jan Tinbergen, League of Nations, Geneva.
– (1940): 'On the Method of Statistical Business-Cycle Research: A Comment', *Economic Journal*, Vol 50, No 197, March, pp 154-56.
Tinbergen, J (1940): 'On a Method of Statistical Business-Cycle Research: A Reply', *Economic Journal,* Vol 50, No 197, March, pp 141-54.

EPW Commentary **October 1, 2005**

Theory or Empirics in Development Economics

Pranab Bardhan

Since I agree with Dilip Mookherjee on the substance of his arguments, much of what I'll say is supplementary or annotative.

In the last two decades there has been a rejuvenation of development economics, and in this the leadership has been provided by a group of empirical researchers mostly in the US. In this period there has been an easier availability of detailed disaggregated, often household-level, data (for example, LSMS data of the World Bank for several countries and census data in many countries, apart from the Indian NSS, ICRISAT and NCAER data). Large memory capacities in our desk computers and sophisticated statistical analysis software have made the handling of gigantic datasets easily manageable, with a speed that still amazes some of us who, when we started our own empirical work, had small Facit machines to be hand-cranked for calculations (including inverting reasonable-sized matrices in the most labour-intensive way), and for large jobs had to wait for days in the queue in the main large-frame computer for the whole campus. Also the influence of labour economics, where applied econometrics had advanced in a major way a few years earlier, has been substantial in allied fields like development. Much more meticulous attention is now paid to finding appropriate identification strategy in our econometric analysis of causal explanations in a way that makes some of the earlier empirical work in development (including my own) look more cavalier. As empirical work has become more respectable in development economics, and sophisticated empirical strategies (including that of random evaluation) explore new horizons, whole armies of graduate students and young faculty are now excited by the field, and nearly 90 per cent of the papers presented in development seminars in the US (including mine at Berkeley) are now mainly, and proudly, empirical.

This is a big change from the days when the brighter students in development would not dirty their hands with the (inevitably messy) data, when the intricate theorems in optimum growth, general equilibrium, imperfect information, or game theory bekoned alluringly. The question that Mookherjee posits suggests that the pendulum may now have swung too much in the other direction. As someone who has dabbled both in theory and empirical work, my inclination is to agree with him and also point, as Mookherjee does, to some of the limitations of the newly

fashionable empirical approach, without doubting its valuable contributions. I shall, however, qualify this with some special remarks on the empirical work on Indian development at the end.

In order to slay the dragons of "endogeneity" and "reverse causality" that threaten the validity of inference in many of the empirical exercises researchers are forever in search of clever "instrumental variables". Once one hits upon a clever "instrument", then the rest is considered relatively secondary, and you crank out your triumphant causal explanation from the two-stage regressions. In this we are often careless about the many pitfalls of instrumental variables, fail to examine the larger meaning of the identifying assumptions, and jump to conclusions. We forget that an instrumental variable even if it has satisfactory statistical properties does not by itself give us an adequate causal explanation or a satisfactory testing of a theory. Let me take the example of an article which is probably the most widely cited in the recent literature on how development depends on institutional quality: Acemoglu, Johnson and Robinson (2001) – AJR analysis. It so happens that I am often on the same side as these authors when it comes to underlining the importance of institutions in development economics, but I cannot go along with the way the profession seems to have lapped up their results as authoritative in quantifying the effect of institutions, and over-interpreted their significance. In their search for an instrumental variable that exogenously affects institutions but not income directly, they imaginatively picked a historical variable, colonial settler mortality, and used it in their first-stage regressions, and the predicted institutions then explained inter-country differences in per capita income in an econometrically cleaner way than has been the case before in this literature. Their presumption is that the mortality rates among early European settlers in a colony determined whether the Europeans decided to install resource-extractive or plundering institutions there, or to settle and build European institutions, like those protecting property rights which help long-run investment and development.

Importance of Institutional Quality

Even ignoring the usual questions about the quality and comparability of data that afflict such cross-country empirical exercises, the fact that some of the early colonies (say in Latin America) were run by Europeans (from Spain and Portugal) at a time when they did not have those property rights institutions quite in place even at home, and the fact that there are many developing countries which largely avoided colonisation (for example, China, Thailand, Mongolia, Ethiopia), this particular exercise has many substantive and methodological problems:
(a) Are property rights institutions the only type of institutions that matter for development? What about participatory and accountability institutions, or institutions of investment coordination? Bardhan (2005) quantifies the effect of some participatory and accountability institutions in explaining the cross-country differences in human development indicators.
(b) As Przeworski (2004) points out, the procedure of instrumenting recent institutions by referring to some old historical fact is flawed because institutions change over time. An instrument for the initial institutions need not be a valid instrument for the current ones. If good institutions are more likely to survive in more affluent countries, then institutional quality today is still endogenous with respect to income.
(c) Albouy (2004) corrects some flaws in the AJR settler mortality measure and shows that when the revised data are used, the AJR analysis suffers from a "weak instrument" problem. (In general in some of the recent use of the instrumental variable approach it is often ignored that a weak instrument can be worse than no instrument). AJR, however, has contested Albouy's corrections.
(d) The disease environment in the 18th or 19th century may be correlated with that today, and the latter affects current incomes directly as well as through its effect on institutions.

An econometrically cleaner method of establishing causality is through random evaluation, which, borrowing a method from medical experiments, has become increasingly popular in development economics, as Mookherjee has noted. This is a very promising development and some researchers have made a creative use of this, including getting involved with NGOs in the field, as participant observer/researcher and influencing their survey design to generate robust

statistical estimates. One of the best examples of work in this genre is Miguel and Kremer (2004), on the impact of deworming drugs among Kenyan school children; this is one of the cleanest empirical assessments of externalities, which development economists always talk about but find hard to measure. But I share some of the general doubts expressed by Mookherjee on the methodology in general. By nature these experiments are much too microscopic, and there is a danger of missing "the forest for the trees". A rush to generalisations from these experimental results will be unwarranted, as they ignore macro- or political-economy or general-equilibrium effects of a programme when they are extended to a larger scale, and the whole is usually more than a sum of the parts. How reliable will it be to generalise the treatment effects of an NGO or government programme, when it is implemented nationwide (as in the case of PROGRESA in Mexico, Grameen Bank in Bangladesh, or self-help groups in micro-finance in India)?

And after all there is more to development economics than more precise programme evaluation or the impact study of a particular intervention. I can see that the latter is particularly important for some administrators and loan givers; I can also see why international financial institutions (like the World Bank) can benefit from the results of these studies on their many projects, which are currently often crudely evaluated. But the task of development economists is beyond writing more effective policy manuals; at least some of them should think about larger structural and conceptual issues. Our improved identifying strategies and controlled experiments have not made us any wiser in deciphering the mechanisms through which certain outcomes are generated (the "why" and the "how") and the social dynamics that are involved, and without these our causal explanations are weak, for all the precision of our tools or the statistical significance of our estimates. In fact, we are sometimes so obsessed with the precision of these tools, that we dismiss potentially insightful exercises that do not pass the standards of our econometric vice squads, and we often let the best be the enemy of the good.

In fact, our preoccupation with accurate quantification often takes us away from the more important causes of a phenomenon and we concentrate on variables that are better measured but may be socially and economically less significant, reminding one of the oft-repeated charge against economists, who look for the missing keys not in the dark place where they lost them but where there is more light. We tend to work with a thin conceptual menu and a large box of precision instruments. In contrast, sociologists and anthropologists spend much less time on honing these instruments, and more on a richer understanding of the processes, relations and dynamics.

Directions for Theoretical Research

In calling for a more balanced portfolio of research Mookherjee has briefly cited some of the directions which theoretical research can fruitfully take. Let me add a few more to his list. We don't know enough about the relationship between factor market imperfections and the social norms, and how this relationship evolves with changing demographic and technological circumstances; how economic processes and community institutions interact; how we can go beyond the existing partial-equilibrium models of oligopoly to generate a viable theory of factor prices and income distribution in a generally oligopolistic economy; an integrated theory of vertical product differentiation, firm heterogeneity, marketing economies of scale and international competition; the nature of transitional dynamics (as opposed to comparative statics) as we move away from a traditional low-level equilibrium; how economic processes get moulded when the state is weak and distant, the legal system is poorly enforced, and there is an "oligopoly" of violence (as opposed to the "monopoly of violence" that Max Weber ascribed to the state); and so on.

Also, in the recent theoretical and experimental work in economics on departures from self-interested maximising behaviour not enough attention has been paid to special behavioural issues that arise in poor countries. For example, the ideas emanating from the growing literature on fairness and reciprocity in individual economic behaviour need to be integrated with the idea of

“moral economy” that historians and sociologists have talked about both with respect to European and Asian cases: for example, Rude (1964) and Thompson (1963) cite cases of peasant jacqueries where peasants in their rage against high prices sometimes looted the granary of a merchant but then paid him what they considered a “fair” price; Scott (1976) has cited cases in colonial Vietnam when peasants rebelling against landlords seized the harvest but paid the landlord what they considered a “fair” crop share. The interesting angle from the anthropological literature on this is that traditional notions of fairness are often asymmetric: (a) in acts of reciprocity or gift exchanges between two parties payments are not supposed to balance, and (b) moral economies often have double standards depending on the domain of the moral community (a peasant who is meticulously fair-minded in his transactions with fellow members of the community has no scruple in cheating people whom he considers outsiders). Our theoretical literature, while beginning to handle normative and weak-willed or time-inconsistent behaviour, is yet ill-equipped to tackle another type of social action which sociologists have pointed attention to: problem-solving interaction with others in which our ends and means co-evolve, with ends discovered and transformed in the process.

The psychological literature on cognitive dissonance and internalisation of severe constraints has obvious relevance to the way the poor behave: as Sen (1984) reminds us, “many of the inequities of the world survive by making allies out of the deprived and the abused”, and as Appadurai (2004) reminds us, the poor may lack “the aspirational resources to contest and alter the conditions of their own poverty”. The constraints they internalise are not just the ones they themselves face, but their parents, peers and neighbours have faced. There is much to draw here from the sociological literature on role models and peer effects and the growing socio-economic literature on neighbourhood effects (mostly relating to poverty traps in US inner-city ghettos). This also relates to the role of group pride, group anxiety and other culturally sanctioned collective processes which deeply affect our preferences (particularly in developing countries suffused with ethnic, caste, or religious affinities), and are yet largely ignored by economists. This has implications for incentives and organisations in labour markets – as Akerlof and Kranton (2005) have emphasised, but also for our theories of collective action (which are dominated by simplistic free-rider presumptions). It raises a larger methodological issue as well. Methodological individualism which undergirds most of economics (including the recent attempts to depart from postulates of self-interested maximisation) may not be an appropriate principle when issues of group dignity and autonomy supercede individual-oriented motivations.

Dearth of Careful Empirical Work in India

Finally, while I may agree that there is too little emphasis on theory compared to the newly fashionable, though highly valuable, empirical exercises, this is mostly in relation to development economics as practised in the west. I have to qualify my agreement in relation to development economics as practised in India, where I believe the old brahminical tradition of high premium on theory still persists. We have a fascination for the esoteric intricacies of a theoretical problem, much less for the sweat, toil and tears involved in grubby empirical work. In particular, there is a dearth of careful empirical work on many vital issues of the Indian economy. While there is a massive amount of largely descriptive empirical work published every year (including in the pages of the EPW), a great deal of it would have improved in explanatory power if more attention were paid to the identification strategies and selection issues that worry econometricians. This also would have stimulated the need for better data collection and for trying out experimental methods in the field, involving real farmers, real workers and real entrepreneurs. There is a great need for more empirical work in India on the structure and practices of enterprises in the vast informal sector (more than 80 per cent of even the non-agricultural labour force is employed there), on interrelationships among firms in a given industry, patterns of industrial evolution with the use of firm-level data, on how our marketing and informal insurance institutions work both in agriculture and outside, work place practices, recruitment mechanisms and wage-setting institutions in industry, trade, and the service sectors, on patterns of technology diffusion, etc. We have a relative abundance of household data on consumption and employment. This has fuelled the

endless debates on measurement of poverty and inequality over the last four decades. Yet to this day we know very little on inter-generational mobility, which is probably the most important aspect of inequality in an extremely hierarchical society like ours, and there are very few attempts at collecting the requisite longitudinal surveys of families. Compared to other countries, we have much less solid empirical work on the necessary restructuring of our pension system or the building of a comprehensive health insurance system outside the formal sector. Fewer good economists are now working on Indian economic history. Just to give one example, there is now a theoretically-informed economic-institutional history of long-distance trade and credit in the Mediterranean and western Europe in the early modern period (for example, the work of Avner Greif). Similar work cries out to be done for India, with its rich history of long-distance trade and credit, of "hundi" and "hawala", and how the caste panchayats and other multilateral reputation mechanisms enforced the rules of conduct. I could go on and on.

Email: bardhan@econ.berkeley.edu

References

Acemoglu, D, S Johnson and J A Robinson (2001): 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, 91(5).
Akerlof, G A and R E Kranton (2005): 'Identity and the Economics of Organisations', *Journal of Economic Perspectives,* 19(1).
Albouy, D (2004): 'The Colonial Origins of Comparative Development: A Reinvestigation of the Data', unpublished, University of California, Berkeley.
Appadurai, A (2004): 'The Capacity to Aspire' in V Rao and M Walton (eds), *Culture and Public Action,* Stanford University Press, Stanford.
Bardhan, P (2005): *Scarcity, Conflicts and Cooperation,* MIT Press, Cambridge, Mass.
Miguel, E and M Kremer (2004): 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', *Econometrica*, 72(1).
Przeworski, A (2004): 'Some Historical, Theoretical, and Methodological Issues in Identifying Effects of Political Institutions', unpublished, New York University, New York, September.
Rude, G F E (1964): *The Crowd in History: A Study of Popular Disturbances in France and England 1730-1848,* Wiley, New York.
Scott, J C (1976): *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*, Yale University Press, New Haven.
Sen, A K (1984): *Resources, Values and Development,* Blackwell, Oxford.
Thompson, E P (1963): *The Making of the English Working Class,* Pantheon Books, New York.

EPW Commentary **October 1, 2005**

Goldilocks Development Economics

Not Too Theoretical, Not Too Empirical, But Watch Out for the Bears!

Ravi Kanbur

In this brief note I would like to set down some of my thoughts on the issues raised in this symposium. My perspective is to see development economics through the lens of mainstream economics. My conclusion is that the balance between theory and empirics is an ongoing process, in economics no less than in development economics. No doubt the pendulum will swing

this way and that, and each swing will bring about its own correction. The balance will tend to be restored. But the really big issues for development economics are also the big issues for economics as a whole – namely, those that arise from our adherence to methodological individualism in a framework of "rational" choice, and the use of overly simplistic economic analysis in policy-making. These issues are neither solely theoretical nor empirical. But they are fundamental to economics and thus to development economics. Hence the title of this note. Goldilocks in the bears' den found the right bowl of porridge for herself – not too hot, not too cold, just right.[1] She found the balance on porridge, but slightly neglected her fundamental predicament.

Theory versus Empirics?

The late David Champernowne, a British mathematical economist of rare distinction, was once asked how much mathematics is necessary in economics. "The amount of mathematics I know", was his answer.[2] There was a time when "mathematics" was associated with "theory" in economics. This is no longer the case. Mathematical techniques are as used in theoretical as in empirical economics, the techniques of mathematical statistics being prominent in the latter. The ground has shifted, since significant proficiency in mathematics and statistics is now the requirement in the best graduate economics degree programmes in the world. Rather, it would seem, the major divide is now between using these techniques in theoretical versus empirical exercises. "The amount of theory I do", might now be Champernowne's apocryphal answer to the analogous question for modern times.

In economics generally, the last 15 years have seen a major shift away from the prominence that was accorded to theory in the preceding decade and a half. One indicator of this is the Clark medal of the American Economic Association, awarded every two years to the most outstanding American economist under the age of 40. Of the last seven awards, only one was awarded to a pure theorist.[3] Of the seven before these, five were awarded to pure theorists.[4] This shift is reflected in the shift in graduate programmes, and appointments in top universities.

Development economics, which has itself become more integrated into mainstream economics in the last 30 years,[5] has also followed this trend. After a long period of concentration on theory, research in development economics entered a strongly empirical phase around 10 to 15 years ago. This coincided with greater and easier availability of data, and of computational power. To take just one example, in the mid-1980s there were only one or two African countries that had nationally representative household sample surveys. Now more than a dozen countries have such surveys, and there are several panel data sets covering households and firms for half a dozen or more countries. The situation was always better in Asia and in Latin America, but there too there have been improvements, sometimes inadvertently so – for example, the opening up of China in the late 1970s and 1980s led to the generation of many micro-data sets. Added to this is the continued improvement, or at least the easier availability, of cross-country data sets for developing countries that are at last nominally comparable – a good example of this is the Deininger-Squire data set on income distribution, which was made available worldwide through the World Bank, allowing a whole industry of cross-country regressions on inequality and growth to develop. The same is true of many other types of macroeconomic data sets that are easily available and easily useable. This is what leads Dilip Mookherjee to say: "Development economics is increasingly becoming an empirical discipline today."

Mookherjee characterises four stages of research: Stage 1 is empirical description, Stage 2 is formulation of theory, Stage 3 is testing, and Stage 4 is prediction and policy evaluation. The focus in the last 15 years has been on Stage 3. Following where mainstream economics has led, development economics these days focuses greatly on testing, in particular the difficulties of inferring causality from correlation. This is the stuff of classical debates, like that between Keynes and Tinbergen, highlighted by Mookherjee. Mookherjee thinks that this focus is now tending to dominate the development discourse, and Bardhan and Basu agree. This is particularly true for

two strands of the empirical literature – that which focuses on the search for instrumental variables, and that which uses experimental design of randomised evaluation. These take us to the frontier of empirical methodology in economics, and according to Banerjee it is this that explains, partly, why development economics is respected within the discipline of economics: “I think the fact that there is no special pleading for empirical work using developing country data is the reason why the brightest and the best among graduate economics students are coming into the field now.”

For both of these strands, it is not the fact of their existence that troubles Bardhan, Basu and Mookherjee, but rather its current preponderance in the journals. This attention to empirics may, according to Mookherjee and Bardhan, crowd out attention to a whole range of interesting theoretical phenomena, such as “…dynamic implications of factor market imperfections for individual agents…, interaction of economic processes with political institutions and social norms...” (Mookherjee), or “…how we can go beyond existing partial-equilibrium models of oligopoly to generate a viable theory of factor prices and income distribution in a generally oligopolistic economy;…the nature of transitional dynamics (as opposed to comparative statics) as we move away from traditional low-level equilibrium; …” (Bardhan).

In addition, and this is a charge particularly levelled against randomised experimental designs, that the very thing that makes inference “clean”, namely, a specific controlled setting, makes any generalisation from this inference problematic. This is the line taken in particular by Basu in his contribution to this symposium. But in his defence of this empirical method, Banerjee sets up the choice between one piece of evidence where the inference is clean, compared to several, from different locations and places, where the inference is problematic, and there is no doubt as to where he comes out: “…even if we have many low quality regressions that say the same thing, there is no sense in which the high quality evidence becomes irrelevant – after all, the same source of bias could be affecting all the low quality results.” Having said this, however, he does of course accept that having many randomised experiments in different locations and at different times is the way to build confidence in their results.

Overall, my own sense is that the protagonists in this debate are not that far apart. Nobody would argue that the use of simple correlation to infer causation should be the modus operandi of empirical development economics. On the other hand, there is a trade-off encountered as the environment generating the data is sufficiently controlled to make clean inference possible – in the extreme, the inference is valid only for that environment and no other. The role of theory in helping to make the best of this trade-off is not disputed either. Fifteen years ago, there was clearly too little empirical work in development economics that met the current frontier standards in establishing inference. Fifteen years on, as the fruits of this empirical excursion have begun to be gathered, there is equally a need for theoretical perspective to make the best use of the empirical results, and to continue working on unresolved theoretical issues. Good sense will surely prevail and we will have both more theory and more good empirical work – a goldilocks solution for development economics.

Watch Out for the Bears

But, in my view, the real problems for development economics, as for economics, lie somewhat at a tangent to the balance between theory and empirics. The “theory” in the theory/empirics balance is none other than modern neoclassical economics, where there is recognition of various types of market and information failures, and outcomes and institutions are understood in terms of these imperfections. This is surely an advance on the general equilibrium theory of 30 years ago, which was not even thought to be of much relevance to development economics. The integration of development economics into mainstream economics has been helped by the direction in which mainstream economic theory has moved, and it has to be said that this move is itself related to the fact that for economists like Amartya Sen and Joseph Stiglitz, standard general equilibrium theory was very far removed from the realities of the developing countries.

However, there remain two basic problems with the current state of development economics – one is in the realm of theory, the other in the realm of policy. In the realm of theory, the basic formulation of individual behaviour, even in "Stiglitzian" models with imperfect information and so on, is one that satisfies the textbook axioms of "rational choice". The behavioural economics revolution in mainstream economics is bringing into question the empirical validity of these axioms, and thus raising questions for those sub-disciplines, like development economics, public economics or labour economics, that are heavily dependent on the framework of standard rational choice. These questions, raised originally by empirical research in behavioural economics, are now being asked anew as similar findings emerge from the empirical development economics. As Banerjee notes in his commentary:[6] "The most intriguing results from empirical research today, as I see it, are not the ones cited by BBM, but results like those of Bertrand-Karlan-Mullainahan-Shafir-Zinman (2004) (the decision to take a loan is at least as influenced by those whose picture is on the offer letter as it is by large differences in the interest rate),…Duflo-Kremer-Robinson (farmers say that they do not buy fertiliser because they have no money, and do not buy fertiliser if it is brought to them while they have money in hand, but will not go and buy it at the local store), etc...It is not just that we do not have a theory within which these results can be interpreted – it is not even clear how we could begin to build that theory." Bardhan, Basu and Mookherjee also highlight the need for new departures from the current centre of gravity of development economics theory when they raise the question of the important role of norms in economic and social behaviour. Bardhan further emphasises the point when he says, "Methodological individualism which undergirds most of economics (including) the recent attempts to depart from postulates of self-interested maximisation) may not be an appropriate principle when issues of group dignity and autonomy supercede individual oriented motivations."

The second major problem is not so much with development economics theory as with development economics policy analysis. No matter how sophisticated our theory and empirics become, it seems that in the debates on great policy issues of the day – on exchange rates, trade policy, labour market policy, deregulation, etc – simple "ECON 101" economics is very important. For many if not most policy analysts, the basic competitive model (and a partial equilibrium one at that) is the workhorse tool – it is the framework that slips most easily into mind when thinking about economic policy. Even some of those who made their names developing the theory of economic policy in a world with imperfections, tend to fall back on the simple policy prescriptions of ECON 101. The justification given is often of a political economy nature, that the more complex policies called for by modern theory (and empirics) are likely to be used and abused by special interest groups. At the very least one can say that this position needs to be tested in theory and in empirics.[7]

Thus, even as we grapple with the perennial question of the right balance between theory and empirics, we should also focus on the fundamentals of the development economics we teach our students, and how policy-makers use (or do not use) the development economics we teach.

Email: sk145@cornell.edu

Notes

1 The term "goldilocks economy" is often used for an economy that is nicely balanced between having too little aggregate demand and having too much.
2 He told this story in his "mathematics for economists" lectures, which I attended at Cambridge.
3 The awardees, in reverse order starting in 2005, were Daron Acemoglu, Steven Levitt, Matthew Rabin, Andrei Shleifer, Kevin M Murphy, David Card and Lawrence Summers.
4 The awardees, in reverse order starting in 1991, were Paul Krugman, David M Kreps, Sanford J Grossman, Jerry A Hausman, James J Heckman, A Michael Spence and Joseph E Stiglitz.
5 For a detailed development of this argument, see Kanbur (2002) and Grusky and Kanbur (2005).
6 See also Kanbur (2003).
7 This argument is developed further in Kanbur (2001, 2002).

References

Grusky, David and Ravi Kanbur (2005): 'The Conceptual Foundations of Poverty and Inequality Measurement' in David Grusky and Ravi Kanbur (eds), *Poverty and Inequality*, Stanford University Press.
Kanbur, Ravi (2001): 'Economic Policy, Distribution and Poverty: The Nature of Disagreements', *World Development.*
– (2002): 'Economics, Social Science and Development', World Development, Vol 30, No 3, pp 477-86.
– (2003): 'Behavioural Development Economics', comments at the conference on Economics for An Imperfect World, in honour of Joseph Stiglitz, Columbia University, October 24-25, 2003, http://www.arts.cornell.edu/poverty/kanbur/Stig-Kanbur.pdf, Accesses, Augus 24, 2005.

Available
EPW on CD
2003 and 2004

New Empirical Development Economics

Remarks on Its Philosophical Foundations

Kaushik Basu

EPW Commentary **October 1, 2005**

I
Introduction

The methodology of scientific research is largely a matter of intuition and knack. As a consequence it is hard to think of formal criteria for evaluating methodology. This must be one reason why the philosophy of science has made so little progress compared to science. The same is true of economics. The subject has made huge gains. But the skirmishes that have been fought about its methodology – for instance, the debate based on Friedman's famous paper (1953) – have done little to enlighten us. Nevertheless, there is value in occasionally pulling ourselves back and asking whether the method that we use to advance knowledge in a particular discipline is right. Dilip Mookherjee's masterly essay, based on a symposium at Cornell University, does precisely that. It is an exercise in introspection and an invitation to practitioners to take stock of the methodology of development economics.

There may be special value in taking up this invitation because development economics has witnessed an upsurge in innovative empirical research. The new empirical development economics[1] is a remarkable achievement. It is not surprising that, once its method was properly understood, there was an explosion of research using its techniques [by, for instance, Esther Duflo, Raghabendra

CURRENT STATISTICS

Macroeconomic Indicators (01 October 2005)

Key Fiscal Indicators for the Centre, States and Combined

Click Here for Weekly Email Content Alert

Chattopadhyay, Edward Miguel, Simon Johnson, Michael Kremer, Daron Acemoglu, James Robinson, Kaivan Munshi, Rohini Pande]. But I have some critiques to offer of this method, which overlap partly with the criticisms put forth by Bardhan (2005) and Mookherjee (2005), and goes beyond to some rather philosophical matters. I shall in this essay concentrate on what, in Dilip Mookherjee's paper, is described as Stage 4 research and a little on his late Stage 3 research.

One reason for the popularity of this new method[2] is the precision of its findings and because, once a result is discovered by this method, we fully understand what it is that has been discovered. Let me consider here one particularly elegant paper written in this mode, namely, Chattopadhyay and Duflo (2004). One of the results reported in this paper is the following: In West Bengal, having a woman head a panchayat has made a difference to what the panchayat does; it leads, for instance, to the better provisioning of water in the village.

In a lot of empirical research, there is the risk that reality may be the reverse of what is being claimed. For example, suppose that in areas with poor water-supply, women (whose traditional job is to fend for the household's water supply) are so preoccupied with the task of fetching water that they cannot participate in local panchayat politics. This could easily create the impression that if women participate in panchayat politics then this leads to the better provision of water. In this case the deduction would be wrong; the causality runs the other way around. What is remarkable about this new empirical development economics is that by using the fact of an exogenous randomisation or the method of carefully-selected instrumental variable it has found a way to get around this problem.

Findings like the above one are extremely valuable; they tell us important things about the way the world is. A popular view, often encountered in newspapers, is that, in a traditional society like India's, having a woman head a panchayat is of no real consequence because she would invariably do her husband's bidding or be bulldozed by the male members of the panchayat. The study by Chattopadhyay and Duflo tells us, compellingly, that that is not so. The new empirical development economics aspires to results of this kind. Where this method applies (and as Mookherjee reminds us, it does not apply everywhere) it is an excellent method to use.

If despite this, I have criticisms of the new empirical development economics, it is clearly not about the method itself but about what we do with it and how we interpret it. And on this there are many popular misconceptions that need to be countered.

Furthermore, it should be stated at the outset that I have some skepticism about all human endeavour to acquire knowledge. Hence, the limitations that I point to here apply much more widely than my focus of attention in this essay, to wit, development economics. I use the new empirical economics as a peg for my argument precisely because it constitutes some of the best research going on in the field.

II
Prediction and Policy

First, it must be recognised that the new method, at least by its own criterion of what constitutes correct inference, does not help us predict the future. Put differently, suppose we apply the strictness of criterion that this new method demands of empirical work, and then ask what we can predict based on this new method? It is important to understand that the answer is “very little”.

Suppose a researcher, studying the effect of aspirin, administers a very low dose of the medicine (150 mg, say) to a random sample of people found walking in the streets of Delhi and finds that, if a person with a headache takes aspirin, he typically (meaning, let us say, in 90 per cent of the cases) benefits from it. Let me call this the “research result” (RR). Now suppose you are asked what would have been the effect of giving 150 mg of aspirin to a random selection of people in Delhi who were lying down with a headache? Based on the RR could you say that they would have benefited from it? The answer clearly is no; you cannot say much, since this sample does not belong to the population on which the original study was done.

Let us now consider how we may use the RR for forecasting. If we next year give the people of Delhi, who are walking around, 150 mg of aspirin, could we expect that those with a headache would benefit from this? Strictly speaking, the answer is no. The people of next year’s Delhi are not the population from which the RR was derived. Also the state variables may be different. The weather may be warmer, there may be more suspended particulate matter in the air, and so on. So to use the RR and to predict the future is like doing a study of the effects of administering aspirin to people who are walking and then presuming that it will apply to those who are lying. If we are fussy about proper randomisation for our study and take the view that we should not accept the wisdom of samples drawn in a biased manner or from the wrong population, we should also take the view that we cannot say anything about the future.

This, in turn, means that we cannot make policy prescriptions, since those are always recommendations for the future. This does not mean that the initial RR is useless. If a person told you that he had (at the same time that the aspirin research was being done in Delhi) selected people randomly from those walking in Delhi and given them 150 mg of aspirin, you would rightly be able to guess that 90 per cent of those with a headache would have benefited from this. The trouble arises with going over from yesterday’s population to tomorrows.

One may try to counter this by arguing that between yesterday and tomorrow there is no fundamental difference and so no reason to expect a relation that was true yesterday to be not true tomorrow. But the difference between yesterday and tomorrow is not just a matter of time. Between yesterday and tomorrow there can be war and pestilence; between yesterday and tomorrow can be 9/11, altering the way world politics is conducted; between yesterday and tomorrow we can have a warmer globe.

One may respond to this further by saying that wars and pestilence do not make a difference to the human physical constitution and so we would expect the aspirin result to carry over from yesteryear to the future. This is perfectly reasonable but in making this argument we are immediately conceding the role of intuition. We are combining our statistical finding with our prior "knowledge" that, for matters of health, knowledge acquired from one population can be carried over to another. We may hesitate to do this about the role of women in panchayats but feel confident about aspirins. This brings us to precisely the point I would endorse: These statistical findings are not useless for prediction but they have to be combined with unscientific intuition for them to be considered useful. We cannot reject the unscientific and claim that our method has predictive power.

Our intuition or our unscientific judgment comes in gradations. Before aspirin is tried on anyone, we may have no faith in aspirin; once we know it has worked on people last year we have more faith in it next year. And the faith grows inductively. For that matter the fact that it has worked on people walking may make us inclined to believe somewhat that it will work on people lying down. But note that there is nothing objective about these beliefs. As I will argue later, what we so often take to be features of the world are actually propensities of the mind.

Let me move on from the subject of prediction and assume (maybe for reasons of intuition just discussed) that time cannot make a difference to the subject matter that is being researched. So a relation found to be valid today will be valid tomorrow – this is an axiom. In that case we can of course predict the future but there remains another fallacy that we can fall into if we are not careful in interpreting the results of the new empirical development economics.

Notice that, if we draw a person at random from among the Japanese who happen to be walking in Delhi,[3] can we then use the above research result and say that he will benefit (meaning with 90 per cent probability) from 150 mg of aspirin in case he has a headache? The answer is no, because this person was not drawn at random from the same population as the original research. He was drawn from the Japanese walkers in Delhi, which is different from the walkers in Delhi.

The stark case where this would be so is if 10 per cent of Delhi's walking population was Japanese, and it was the case that 150 mg of aspirin helped a walking person if and only if he or she was not Japanese. In this case the RR would be true and, at the same time, a randomly-chosen Japanese person would not benefit from aspirin. This in turn leads us to the following troublesome question. Suppose you know of the RR (and this is guaranteed to hold over time). Now you are walking down Janpath in New Delhi with a headache, and wondering if you will benefit from 150 mg of aspirin. The answer depends on whether you can be thought of as a random draw of a person from the walking population of Delhi. It seems reasonable to me to treat the answer to be no. For one, you could be a Japanese in which case you are drawn from the population of Japanese walkers

in Delhi. Hence, you cannot draw any policy lesson for yourself from the research result. This is a rather worrying predicament. It means that whenever I want to use a research finding (based on proper controlled experiments) for my own treatment, strictly speaking, there is little reason for me to have faith in the result, since I am not a random draw from the population.

Am I using too demanding a criterion for what constitutes knowledge? Maybe, but since this new method in development economics is itself based on a very strict criterion of what is statistically correct, those who view this kind of correctness as essential should go along with my skepticism.

III
Knowledge and Evolution

The skepticism about the acquisition of knowledge expressed in the above section is troublesome. Despite my own inclination towards skepticism, I am aware that one must not have an unbending adherence to it (that would be a contradiction anyway for a skeptic). One possible mistake that both the skeptic and the practitioner of the new development economics have to guard against is that of denying that there may be other modes of acquiring knowledge.

To understand this, consider the number of things that we learn from poorly-controlled experiments or, for that matter, no experiments. A growing child learns soon enough that a frown implies displeasure and a smile implies approval that a slap hurts and a massage soothes (especially neck massage), that when a person cries she is sad and when she laughs she is happy.

Suppose this child's father stops the child each time she makes an inference, by asking whether she was sure that she was making the deduction from a proper random sample and not merely from her experience with those she happens to bump into in her everyday life. And suppose he asks her to discard any knowledge not picked up from properly randomised experiments. Surely this child will turn out to be a very ill-informed adult. The fact of the matter is that the knowledge that we human beings carry in our heads is disproportionately from wrongly conducted experiments and from biased samples. The knowledge that we have from scientifically-conducted studies – things like, 80 mg of aspirin a day can cut the risk of a heart attack by half and oatmeal reduces cholesterol – are a tiny fraction of what we know.

This is indeed a puzzle: How do we know so much given the atrociously-biased methods we use through life to collect information? Of course, we make errors. My son, whose first experience of the world outside of India was Belgium, for a long time after our return used to point to any white person he saw in Delhi and say confidently, "Belgian". It took considerable effort on my part to persuade him of the sampling bias that underlay his deduction.

There are three responses we can have to this puzzle. One is to try

to show how, even if each person uses a biased sample, by the act of pooling our individual information, as we human beings do, the biases tend to cancel out and for the large part correct themselves. This would be an interesting research agenda in probability and information theory (even though it would still leave open the question of how we can predict about tomorrow from our information of yesterday).

But if this theoretical exercise turns out to be futile (and till such a result is proved it seems reasonable to proceed as if it were not true), then there are two possible positions that we can take. One is to claim that we human beings actually do not know much. Much of our knowledge is chimera – a mere illusion of knowledge. Many religious traditions and also some irreligious philosophers take such a view. There is a long Greek tradition of this.[4] Historically, the most famous is the philosopher, Pyrrho (4th century BC). Pyrrho did not write down any of his philosophy because he was skeptical about its value (though, of course, he could have been equally skeptical about the value of not writing and written a lot, like Bertrand Russell, also a skeptic,[5] did). It is believed that he went with Alexander's army to India and returned humbled because in India he had met "sadhus" who not only did not write but did not even speak. Legend has it that he heard one of his teachers asking for help, having fallen into a ditch; but he walked away calmly because he could not be sure that the teacher would be better off outside the ditch than in it. The teacher in this case happened to be Anaxarchus, a philosopher who held many similar views to those of Pyrrho. After he was heaved out of the ditch by some others and safe, it is believed that his greatest praise was for Pyrrho whom he had seen walk past with complete sangfroid [Diogenes Laertius 1925].

Maybe it was to counter these extreme versions that the later Greek philosopher, Carneades (2nd Century BC), stressed that, behaviourally, a skeptic need not be any different from a non-skeptic.[6] It may however be recalled that Carneades himself caused some comical problems by arguing one day in favour of justice and another day against it, since he felt committed to neither view. In fact, the Greek physician and philosopher, Sextus Empiricus (2nd century AD), took the view that the main consequence of skepticism was the tranquillity of mind achieved from resigning oneself to the futility of the quest for knowledge.

A second position, which is not incompatible with the above one, is to take an evolutionary view of knowledge [Basu 2000]. This is to admit that we do not know how we know things but, if it is the case that a person's knowledge or (at a more meta level) the facility for acquiring knowledge is inheritable, then the people who have wrong beliefs or knowledge about the way the world works (those, for instance, who think that the frown on the face of the person approaching them menacingly, knife in hand, indicates friendliness) would perish in the long run. So the people we see around know that an apple released in the air will fall downwards and know that the knife being brought down on a person will kill him. The fact of these people being around means that they and their ancestors have survived the weeding

process of evolution. The ones who knew the wrong things or did not have the facility for learning correctly from nature are just not around anymore. By this theory, there is no right way to acquire knowledge. Nature is too idiosyncratic for that. But some minds are synchronised with our idiosyncratic nature and some minds are not. The people we see around us, by virtue of the fact that they exist, have minds that are in synchrony with nature.

My own belief about the puzzle of knowledge lies somewhere between the skeptical and the evolutionary claims. The trouble with this is that unlike others, who have strong views on what is the right method, I do not have any. I therefore find it difficult to take a clear position on debates like whether we need more theory or more empirics. I have some faith in our intuition. Two correctly done empirical results may have the property that one resonates with our intuition – we simply feel that if it was true in the past it has a reasonable chance of being true in the future – and the other does not. My inclination would be to go along with the intuition (while admitting that intuitions often go haywire). And the same is true with theoretically-derived results. Some feel right; some do not. I would be tempted to give in to the feeling. Hence, the issue is not between theory and empirics.[7] We need to do both as correctly as we can and then use our intuition to select the ones we want to live by and base our policy recommendations on those and, most importantly, keep our fingers crossed behind our backs, when doing so.

IV
A Comment on Causality

Economists have a propensity to find the lack of evidence of causality in other people's research. We complain about how other people's empirical papers demonstrate correlation but not causality. Such language is fine, as long as we realise that there is no real way for demonstrating causality. Indeed, there is no reason to believe that there is anything objective in nature called causality.

It may sound baffling at first sight but it is much more robust to maintain that causality lies in the eyes of the beholder. That is, we human beings have it hardwired in us to think in terms of causality. This can be a useful feature of our mind. It allows us to be more sure-footed than we would be if we could not live as if causality existed.[8] And, as we saw in the last section, there is reason to believe that what our mind takes to be causal is reasonably dependable because it has come to be reasonably well synchronised with the way nature actually works, thanks to thousands of years of human evolution. Hence, it is useful for us to feel that a particular relation is causal (and hence dependable in the future) and not merely a matter of correlation. But even on this we can make large mistakes. I will end on this cautionary note by constructing an example of how we can be led astray in our interpretation of probabilities.[9]

Suppose that there was a ritual in the world whereby in each city the mayor tosses a coin 20 times at the start of each millennium. So the last time this was done was on January 1, 2000. Let us suppose in

Washington a researcher wanted to see if there was reason to believe that there was a particular bias in coin tosses in Washington and he discovered that in fact all the 20 tosses had yielded heads. He would be tempted to publish a paper entitled, “Heads Bias in Washington,” and may have speculated if having too many senators or too many lobbyists caused this. In case he had no experience of the rest of the world but had experience of coin tosses by others in Washington, he may have thought he had made the discovery of how the city’s mayor is head prone. It is indeed very surprising to get 20 heads in a row. It does seem to say something to us.

Now return to the problem and think of the larger picture. Given that this ritual is followed in thousands and thousands of cities, the probability is very high that a sequence of 20 heads will occur somewhere. So the fact that this has occurred somewhere is of no interest whatsoever. It reveals nothing beyond what we already knew from our rudimentary knowledge of probabilities.

Let me digress for a moment to deal with a possible technical objection. A critic may say that before publishing the paper the author should have collected more information so as to decide if what happened was just pure chance. He may for instance collect data on another 20 tosses of the coin in Washington. But surely that will not change anything because we can pretend that this has already been done. It could, for instance, be that he had first collected information on the mayor’s first 10 tosses and, seeing that all were heads, asked the mayoral office to send him the data on the next 10 tosses. He was then astounded to find that the next 10 were also heads and therefore wrote the paper. Therefore, we can think of larger and larger number of tosses. As long as there are more and more cities where the experiment can be done, we can construct the same, logically-equivalent story.

To return to the example, what we have are two propositions. The Washington researcher’s feeling that he is onto something interesting seems justified. The outcome that he has seen has too small a probability if the probability of each coin toss by the mayor of Washington landing a head is half. So he is right in concluding that the probability of a head is much more than half. On the other hand, we also know from our larger global perspective that there is actually no special information in what happened in Washington. A string of successive heads occurring somewhere is very high, even if each coin toss has a probability half of getting a head.

The analogue of this problem is that researchers the world over are studying different phenomena. They publish only what seems unexpected. Since the expected does not get published, we never get the larger global picture (like the data on coin tosses from across the world) and so think we have stumbled upon knowledge when, in fact, we have not.

Am I making a mistake somewhere?

Email: kb40@cornell.edu

Notes

[I am grateful to Talia Bar and John Roemer for reading the essay, making helpful comments and criticisms and to Ted O'Donaghue for a very useful suggestion.]

1 Like some of the most interesting concepts in life (consider "expressionist art" or "Victorian manners") "new empirical development economics" is easier to identify than to define. I will, broadly, take it to refer to the recent research which uses controlled randomisations or carefully selected instrumental variables to reveal causal links between economic variables that are of particular interest to developing countries.
2 I refer to this as the "new method" given its relative newness in development economics. The method itself is not new – it has been used for some time in epidemiology, for instance.
3 For those planning to travel to Delhi, I should clarify that this gives a somewhat exaggerated impression of Delhi's cosmopolitanism.
4 An interesting summary account occurs in Long (1998).
5 Russell was however troubled by the fact that conventional skepticism was not just a philosophy of doubt, but what it never should have been – "dogmatic doubt".
6 "We Sceptics follow in practice the way of the world, but without holding any opinion about it" [Bevan 1950, p 52]. This, interestingly, amounts to a critique of behaviourism. But I take behaviourism to be an easy target, summed up (no doubt in somewhat of a caricature) in the observation, allegedly made by Bertrand Russell, that there is no way to tell the difference between a mathematician asleep and a mathematician at work.
7 The limitations that I discuss in this paper for empiricism should not be read as endorsement of theory as the instrument of choice for understanding the way the economy works. Theory can help us sort certain deductive complications but may not be able to do much more [see Appendix to Basu 2000]. For an excellent and persuasive essay on skepticism in the context of theory, see Rubinstein (2004).
8 The fact that human beings tend to pick up patterns beyond what is actually there has been documented in psychological experiments [Tversky and Kahneman 1971]. They report on several interesting findings. For instance, when people are asked to guess what a random sequence of coin tosses will yield, they produce sequences in which heads are closer to 50 per cent than turns out to be the case in reality. This is especially true for short sequences. It could however be argued that this tendency on the part of human beings may have evolutionary survival value.
9 The philosophical foundations of probability have a history of confounding not just philosophers and statisticians but also economists, most famously John Maynard Keynes. Indeed, it may be Keynes's early encounter with probability theory that led to some of his strong views on the empirical method that Mookherjee discusses in his paper.

References

Acemoglu, D, S Johnson and J A Robinson (2001): 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, Vol 91.
Bardhan, P (2005): 'Theory or Empirics in Development Economics', *Economic and Political Weekly,* this issue.
Basu, K (2000): *Prelude to Political Economy: A Study of the Social and Economic Foundations of Economics,* Oxford University Press, New York.
Bevan, E (1950): *Later Greek Religion,* Beacon Press, Boston.
Chattopadhyay, R and E Duflo (2004): 'Women as Policy-Makers: Evidence from a Randomised Policy Experiment in India', Econometrica, Vol 5, 1409-44.
Diogenes Laertius (1925): *Lives of Eminent Philosophers*, Volume II, translated by R D Hicks, Harvard University Press, Cambridge, MA. (The original Greek text is usually dated to the 3rd century AD.)
Duflo, E and R Pande (2005): 'Dams,' mimeo: Massachusetts Institute of Technology.
Friedman, M (1953): 'The Methodology of Positive Economics' in M Friedman (ed), *Essays in Positive Economics,* Chicago University Press, Chicago.
Long, A A (1998): 'Hellenistic Philosophy' in R H Popkin (ed), *The Columbia History of Western Philosophy,* Columbia University Press, New York.
Miguel, E and M Kremer (2004): 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', Econometrica, Vol 72.
Mookherjee, D (2005): 'Is There Too Little Theory in Development Economics Today?', *Economic and Political Weekly*, this issue.
Munshi, K (2003): 'Networks in the Modern Economy: Mexican Migrants in the US Labour Markets', *Quarterly Journal of Economics,* Vol 118.
Rubinstein, A (2004): 'Dilemmas of an Economic Theorist', mimeo: Tel Aviv University, http://arielrubinstein.tau.ac.il/papers/Pa.pdf
Tversky, A and D Kahneman (1971): 'Belief in the Law of Small Numbers', *Psychological Bulletin*, Vol 2.

Available
EPW on CD
2003 and 2004

'New Development Economics' and the Challenge to Theory

Abhijit V Banerjee

EPW Commentary **October 1, 2005**

New Development Economics?

If we take the list of papers cited by BBM (Bardhan, Basu and Mookherjee, all in this issue) as representative of what they call "New Development Economics", what is striking is how old-fashioned it all sounds. The topics, for the most part, are the familiar ones: education, health, credit, technology, land, migration. Even institutions, which sound a bit more novel, were being discussed and debated by North and his colleagues more than 20 years ago. It is true that agriculture is a bit less in fashion now than it was 25 years ago, and there is, to my taste, a bit too much about human capital, but relative to many other fields (think of macroeconomics) the picture is one of remarkable stability.

In terms of empirical methodology, while there is heightened emphasis on running well-identified regressions, the basic concern with distinguishing causation from mere correlation, that it reflects, obviously goes back a long way. The seemingly unending debate among development economists of the 1960s and 1970s about whether bigger farms were less productive (the so-called size-productivity debate), was ultimately a debate about causation – about whether it was really size that was doing the damage or something associated with size. What has changed since then is that development economists have now embraced all the strategies that have been developed in the last two decades, mostly by labour economists, for dealing with identification issues – difference-in-difference, regression discontinuity designs, randomised experiments, etc – and have access to the kind of data that is needed to make use of these techniques.

In other words, these papers reflect, more than anything else, the mainstreaming of empirical work in development economics. The field, as Mookherjee explicitly recognises, has matured.

I
In Defence of Empirical Method

CURRENT STATISTICS

Macroeconomic Indicators (01 October 2005)

Key Fiscal Indicators for the Centre, States and Combined

Click Here for Weekly Email Content Alert

BBM share a feeling that in the process of maturing things have gone a bit awry. Their main complaint, though none of them quite puts it this way, is that empirical standards are now unrealistically high, with the consequence that the top journals are now filled with well-identified but uninteresting papers.

To some extent this has to be right. To accept something that obviously does not meet the standards of the profession because the subject is interesting, editors have to exercise judgment, and exercising judgment is never easy, especially once you recognise that this will expose you to the charge of favouring certain people and/or ideas. There are certainly interesting papers that shed important light onto questions we know little about that do not get published because they do not quite make it on some methodological imperative.

Would I therefore rather that we had not come to a point where development economists are at the cutting edge of empirical techniques? Absolutely not! I think the fact that there is no special pleading for empirical work using developing country data is the reason why so many of the brightest and the best among economics graduate students are coming into the field now. When I was a graduate student, everyone always said that development economics was very important but no one actually did it: The problem was that if you were a top student you wanted to be where people appreciate your mastery of everything that was clever and new (I, for example, did theory, which was very much in fashion then). Now it is development economics that is in fashion.

As far as I am concerned this alone is enough compensation for the sacrifices that we have to make to the harsh god of identification. But I am also convinced that the danger of unwarranted rejections is nowhere as serious as the danger of publishing spurious results. As I have argued elsewhere, there is no dearth of ideas in the world of development practitioners today. What is missing is discrimination: There is no rigorous process by which bad ideas get dropped and good ideas are identified and made better.

A World Bank publication from few years ago[1] provides an excellent example of the attitude towards evidence in the development community outside academia. The *Sourcebook* is meant to be a catalogue of what, according to the bank, are the right strategies for poverty reduction. These are also, we presume, strategies into which the bank is prepared to put its money. It provides a very long list of recommended projects, which include: computer kiosks in villages; cell phones for rent in rural areas; scholarships targeted toward girls who go to secondary school; schooling voucher programmes for poor children; joint forest management programmes; water users' groups; citizen report cards for public services; participatory poverty assessments; internet access for tiny firms; land titling; legal reform; micro-credit based on group lending; and many, many, others.

While many of these are surely good ideas, the book does not tell us how we know that they work. Indeed, one memorable example

makes clear that this is not a primary concern of the authors. The book describes a programme called the *Gyandoot* programme which is based in Madhya Pradesh in India and provides computer kiosks in rural areas. The *Sourcebook* acknowledges that this project was hit hard by lack of electricity and poor connectivity and that "currently only a few of the kiosks have proved to be commercially viable". It then goes on to say, entirely without irony, "following the success of the initiative…" (p 80).

The most useful thing a development economist can do in this environment is to stand up for hard evidence and to be demanding about the kind of empirical work that gets into the better journals. Regressions published in the top places have a way of filtering into policy conversations – the power of statistics combined with the imprimatur of a top journal can be irresistible. A recent example is the work by Burnside and Dollar on the impact of aid on growth: In a paper published in the *American Economic Review* (2000),[2] they showed results suggesting that aid does help growth but only when it is given to countries that pursue the "right" policies. This is, of course, also the idea behind president Bush's Millennium Challenge Account (MCA), and while I do not have any direct evidence of a causal connection between the two, the Burnside-Dollar work is frequently cited as a part of case for the MCA (including by the United States Congress Research Services). Yet there are multiple reasons why we (and especially the editors of the AER) should have been more skeptical of the Burnside and Dollar results. First, because policies are not exogenous: They reflect other things that are right about the country. Second, aid is not exogenous either – countries with bad policies that received a lot of aid may have other problems (for example, dictatorships that are aligned with the US tended to get a lot of aid in the old days), which make it hard to compare them with other countries. In any case, a later paper by Easterly, Levine and Roodman (2003) showed that the result in the Burnside and Dollar paper disappears when they update the data.[3] But the MCA is here to stay.[4]

But how do we know that well-identified regressions are any better? For one, as Basu points out, there is the very real risk of reporting bias: To get published it helps to have the kind of results that the editors or referees want. Basu is worried that this favours papers with unexpected results, but it could also favour results of a particular political bent.

Ill -identified Regressions

Note, however, that this is potentially just as true of ill-identified regressions as it is of well-identified ones. Indeed, the fact that papers these days are getting published because they have a clever identification strategy, something that Mookherjee and Bardhan deplore, has the advantage of reducing the emphasis on results. The main contribution of a number of the papers that BBM cite as examples of new development economics (Acemoglu, Johnson and Robinson (2001);[5] Munshi (2003);[6] etc) was to provide credible evidence for what many others before them have claimed

("institutions matter", "social networks are important", etc).

That being said, I think there is a widely shared feeling that we need to do more about this problem. In particular, I support an initiative to set up a website where every randomised evaluation has to be registered before it gets launched. At the time of registration, the principals of the study will be required to list the outcomes of interest and their predictions regarding each of them (just the direction, not the magnitudes), as well as a reasonable study completion date. They will then be required to report whatever results they have, on or before the announced date. This would guarantee that all studies (including the ones that fail and the ones that have boring results) get recorded, that the authors cannot decide ex post to emphasise interesting outcomes that they never thought of when they started the work, and that the experiment does not continue until it yields the desired result. Doing something like this for non-experimental work would be harder: While experiments tend to be very visible, there is no way to know all the failed regressions that get run – unless the authors want to report them. One thing that might help is to set up a web-space where people can report their failed regressions (just a paragraph . . . we ran this regression in this data set and found the wrong answer or no answer), and then editors can require all future published authors on the same subject to cite these failed studies. In the end, however, this remains another point in favour of randomised experiments over more traditional empirical exercises.

On the other hand, it is true, as Bardhan points out, that randomised evaluations typically give us only the short-term impact of a particular intervention. However, the main reason for this is that the authors of the study want to be able to report some results without having to wait forever, though in many there is nothing, in principle, to stop them from continuing to follow the treatment and control groups after the first round results have been reported. For example, consider Miguel-Kremer's work on deworming cited by Basu.[7] In that study, they randomly assigned a programme of giving free deworming medicine to children across a set of schools in Busia District in Western Kenya and found that the children in the schools where the medicine was given came to school much more regularly. While that paper is published, the study continues: The goal is to discover whether the fact that they came to school more often helps these children in their future life. To do so they will need to follow the children in both the treatment and control groups as they grow older. This is possible because even though the control schools may now be treated, the children who graduated from those schools before the medicine was introduced there continue to be a valid control group for the corresponding cohorts in the treatment schools.

The point is that there is no necessary conflict between the goals of extending the treatment to the control group (assuming it works) and studying the long-term impact of the treatment. There are cases where this will not work – these are programmes where literally every group will end up treated once the programme is expanded – but these are more the exception than the rule. The bigger constraint, it seems to me, is the sheer effort that goes into following people into

the future, but if the stakes are high enough, it will be done.

Among the various concerns raised by BBM, the two most important are what I would call the problem of scope and the problem of size. The problem of scope comes from the fact, emphasised by Bardhan and Mookherjee, that most well-identified empirical exercises tend to be relatively localised: This is what makes it possible to rule out other conflating factors (the extreme case of this is a randomised experiment where we deliberately restrict the domain in order to have full control over the programme placement). This, they suggest, compromises the external validity of these results – how can we draw general lessons from something so limited?

Basu takes this argument a step further. The external validity of a piece of empirical research, however carefully done, he argues, cannot be entirely derived from something internal to that work. To apply these results, we need to have a theory that helps us decide whether a particular location (in space, or Basu emphasises, in time) is sufficiently similar to the location from where the results came. And in most cases this theory would not come close to meeting the standards of evidence that development economics today aspires to, and even if it did, the problem of external validity would remain, just shifted back a layer.

This is obviously correct. Indeed it seems to me to be a version of David Hume's justly famous demonstration of the lack of a rational basis for induction. But at this point Basu makes a curious leap into what reads like radical scepticism. As he puts it, "Two correctly done empirical results may have the property that one resonates with our intuition – we simply feel that if it was true in the past it has a reasonable chance of being true in the future – and the other does not. My inclination would be to go along with the intuition (while admitting that intuitions often go haywire)." This seems to imply (by continuity) that if one empirical conclusion accorded less well with his intuition than another, but was somewhat more "correctly done", would he still go with the latter? In other words, he denies the primacy of empirical methods over his intuition.

I must confess that I do not understand this position. To take an example, one of the early randomised experiments in Busia district of Kenya looked at the effect of having flipcharts in the classroom on the performance of children.[8] It found that the children in the schools that were randomly chosen to get a flipchart do not do any better. The study then looks at what it would have found if it had tried to answer the same question based on running a retrospective cross section regression using data from other schools in the same part of Kenya. In other words, they ran a simple regression of the test scores on whether the school had flipcharts in this other sample of schools. They found that having flipchart raises student performance about 0.2 standard deviations.

My inclination, faced with this evidence, is to assume that the experiment gave more or less the right answer for what would happen in an average school. I start from the fact that the experiment

was correctly done: The experimental result is therefore the right conclusion for those particular schools. The question then is whether this also the correct result for other schools of this type in the area, including the ones in the retrospective study. To believe this, I would need to believe: (a) that the cross section study is biased; and, (b) that the treatment effect does not vary across the two populations.

Say that Basu, on the other hand, believes that the retrospective result is the correct generaliseable prediction because it fits better with his prior intuition (this is entirely hypothetical and almost surely unfair to him, but it helps to set out the structure of our disagreement). He therefore believes that: (a) the retrospective result is unbiased; and, (b) that the treatment effect is not the same in these two populations.

These two sets of beliefs (his and mine) are, however, not symmetrical. The fact that the treatment effect varies in this way is not something that he has any independent reason to believe: The study which (I assume) summarises everything either of us knows about the area, suggests that the populations are in fact quite similar. Or to put it differently, the only reason to suppose that the treatment effect is higher in one of these populations than the other is that this is what squares with his prior – this was not something that he learnt from the data generated by the retrospective study or any other study.

Standard Observation

My suspicion of the results from the retrospective study, on the other hand, stems from the standard observation, confirmed by any number of studies, that schools in developing countries that have more school inputs tend to be schools where parents are richer and/or care more about education and hence are schools where the children would have done better even without the extra inputs. I worry that the fact that children in schools with flipcharts do better than children in schools without flipcharts could be partly a result of this. In other words, I favour the experimental result mainly because I have an independent reason to suspect that the retrospective result is biased upwards.

Moreover, my assumption that the treatment effect is the same in the two populations, while not justified by any direct evidence, would seem to be the natural default assumption in the absence of any information to the contrary. It is clearly the way we are designed to function in the world – when we meet a new child our first instinct is to treat him like all the other children we know – though I would not know whether it is evolution that has made us so (as Basu would have it), or our own experience with what is sometimes called the “uniformity of nature”.

None of this, however, tells us exactly what to do when we expect there to be systematic differences between the two populations. If our only really reliable evidence was from India but we were interested in what might happen in Kenya, it probably does make sense to look at

the available (low quality) evidence from East Africa. Moreover, if the two types of evidence disagree, we might even decide to put a substantial amount of weight on the less reliable evidence, if it turns out that it fits better with our prior beliefs. Nevertheless, there remains an essential asymmetry between the two: The well-identified regression does give us the "correct" estimate for at least one population, while the other may not be right for anyone. For this reason, even if we have many low quality regressions that say the same thing, there is no sense in which the high quality evidence becomes irrelevant – after all, the same source of bias could be afflicting all the low quality results. The evidence remains anchored by that one high quality result.

That being said, the only way to build trust in experimental and quasi-experimental results is to replicate them in several different locations. One of the strengths of the Miguel-Kremer work on deworming cited by Basu,[9] is that it has now been reproduced in India (the original study was in Kenya) with very similar results.[10] Moreover, Bleakley (2004),[11] using quasi-experimental methods to estimate the effect of deworming in the US. South also finds a structural estimate very similar to what these two other studies find. The randomised evaluation of the Balsakhi remedial education programme in India was carried out simultaneously in two different locations (Mumbai and Vadodara) by two different implementing teams for exactly this reason, and, reassuringly, found more or less the same results in two locations.[12] These results suggest that extrapolating these results across large distances may not always be as bad as it might seem, but clearly many more replications need to happen.

Importance of Size

The problem of size arises because randomised experiments in particular tend to be relatively small scale, partly due to practical and financial constraints, but also in part because we want the experimental results before scaling up the programme. Other careful micro studies are less constrained, but the need to know exactly what is going on, which underlies good identification strategies, tends to limit the scope of these studies as well. Size is important for two reasons: One, which applies only to randomised evaluations, is that small programmes are easy to monitor carefully and therefore are not run like real programmes. While this is a concern, many of the more recent randomised evaluations are only small relative to nationwide programmes: The Balsakhi and deworming studies mentioned above, for example, each involved collecting data from tens of thousands of children and a number of other recent studies are even bigger. Moreover there is a lot of emphasis on making sure that the programmes that get evaluated are developed enough to be modular, in the sense that what gets implemented at every location is described by a simple and common protocol. This makes it easier to limit the involvement of the programme's sponsors and the evaluators in the implementation process.

Size is also an issue because of what are often misleadingly called general equilibrium effects: Educating some more children in

Vadodara or Mumbai is not the same as raising educational standards in the country as a whole – presumably if the number of educated people in the country goes up enough all sorts of other things will change. Wages paid to educated people will fall; the political clout of educated people will rise; others might see them and also start demanding more education; etc.

This is a very real problem, though obviously the import we need to give it depends on the question we are asking. If we are interested in how to get children to come to school, we might be okay in ignoring the effect of increased enrolment on wages 15 years hence, since parents probably do not think that far ahead. On the other hand, if we want to figure out how much we add to the future earnings of these children by making them go to school, it would be silly to not worry about what would happen to wages. Even the most committed experimentalist would want to do something.

II
The Challenge to Theory

This is where, Mookherjee argues, we ought to be using more theory. Theory helps because what we really need is to make an assumption, and theory tells us what the right assumption would be. For example, we know from market equilibrium theory that the one number we need in order to assess the effect of the increase in education on wages is the elasticity of demand for skilled labour. Armed with this knowledge, we could go looking for an estimate of the elasticity and having found one (hopefully from an experiment or a quasi-experiment), could use it to calculate the true benefits from getting children into school.

Theory can also help us in solving the problem of scope. The conventional structural approach, which Mookherjee supports, involves fitting a model with a small number of unobserved parameters to the observed programme effects, and then using this model to make out-of-sample predictions.[13] Conceptually, given that parameters of the structural model are estimated using the fact that the programme effect is different for people with different observables characteristics, this is not that different from a naïve approach, where we allow the programme effect to depend on observables, and then extrapolate the results using what we know about the same observables in the new population. The advantage of the structural approach is that we have a theory to guide us about the kind of parametric restrictions we want to place on the data rather than having to guess at it. The disadvantage is that the theory may be wrong. The fallout of the behavioural economics revolution in economics is that we are no longer particularly sure of what the right theory ought to look like, especially inasmuch as decision problems are concerned. In particular, we are no longer secure in the presumption that utility functions and cost functions are somehow more stable and more universal than behavioural rules. Of course, as Mookherjee notes, we could also structurally estimate behaviour rules, but for that we would need a new body of theory. In the meanwhile it is not clear that using the existing theory always

dominates simply assuming an ad hoc empirical specification, but perhaps it is best to use the two approaches in tandem, using one to check on the other.[14]

The bigger challenge to theory however comes from a different direction. The most important role of theory in development economics, and indeed in all the rest of economics as well, is to help us understand what are the right questions. The formulation of a testable hypothesis is only the final stage of this process, and one that is often left for the empirical researchers to do, since it often depends on the exact nature of the data. What is prior to that, and in some ways, even more important, is the ability of theory to locate the empirical results within a broader intellectual context and make us see why we ought to care. In this sense, a lot of the best empirical work of the last decade or so can be seen as a response to a body of theory developed in the previous decade that made us really understand the implications of living in a world where neither markets nor governments work perfectly.

What is unusual about the state of development economics today is not that there is too little theory, but that theory has lost its position at the vanguard: New questions are being asked by empirical researchers, but, for the most part, they are not coming from a prior body of worked-out theory. The most intriguing results from empirical research today, as I see it, are not the ones cited by BBM, but results like those of Bertrand-Karlan-Mullainathan-Shafir-Zinman (2004) (the decision to take a loan is at least as influenced by whose picture is on the offer letter as it is by large differences in the interest rate),[15] Karlan-Zinman (2004) (moral hazard in credit markets is a problem for men and not women, while for adverse selection it is the other way around),[16] Duflo-Kremer-Robinson (farmers say that they do not buy fertiliser because they have no money, and do buy fertiliser if it is brought to them while they have money in hand, but will not go and buy it at the local store),[17] etc. It is not just that we do not have a theory within which these results can be interpreted – it is not even clear how we would begin to build that theory.[18] Indeed, the same goes for some less exotic results like Michael Kremer's finding that teachers teach better when students are given an incentive to do well (2004),[19] but not when teachers themselves are rewarded if their students do well (2001).[20]

The point is not that these phenomena are more complex than what we are used to and therefore need a more sophisticated theory. The problem is precisely that the correct explanation may not be particularly interesting or illuminating: Perhaps the borrowers in Bertrand-Karlan-Mullainathan just liked the picture and responded to it. Maybe the farmers in Duflo-Kremer-Robinson had never thought of the possibility of going to the store when they had the money (which is a few weeks before they need the fertiliser). Could it not simply be that the students in Kremer's study are better at getting the teacher to do what they want than the teachers are in getting cooperation from the students?

The challenge is to turn these explanations into a theory – something

that would have implications for other things that we care about and pose new questions for empirical research. I believe that it can be done, but to do this we need to give up trying to defend the existing theory (which has been incredibly successful in many ways) against an onslaught of seemingly random results that are coming out of the field experiments. In Haavelmo's terms (as described by Mookherjee in the present issue) we are back at stage 1: We are gathering impressions that would eventually allow us to build a set of new theories that, one day, will define a "newer" development economics.

Email: banerjee@mit.edu

Notes

1 Deepa Narayan,, Empowerment and Poverty Reduction: A Sourcebook, World Bank, Washington, DC, 2002.
2 Craig Burnside and David Dollar, 'Aid, Policies, and Growth', American Economic Review, September 2000, 90(4), pp 847-68.
3 William Easterly, Ross Levine and David Roodman, 'New Data, New Doubts: A Comment on Burnside and Dollar's 'Aids, Policies and Growth (2000)' ', mimeo Department of Economics, New York University.
4 Indeed, given all the identification problems with running this regression, there is no reason to presume that either result is actually correct. In other words, the real problem is that we have not learnt anything from the evidence. It therefore remains entirely possible that the MCA will end up doing much good in the world.
5 D Acemoglu, S Johnson and J A Robinson, 'The Colonial Origins of Comparative Development: An Empirical Investigation', American Economic Review, 91(5), 2001.
6 K Munshi (2003), 'Networks in the Modern Economy: Mexican Migrants in the US Labour Markets', Quarterly Journal of Economics, Vol 118.
7 E Miguel and M Kremer (2004), 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', Econometrica, Vol 72 (1): 159-218.
8 Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz (2004), 'Retrospective vs Prospective Analyses of School Inputs: The Case of Flipcharts in Kenya', Journal of Development Economics, 74: 251-68.
9 E Miguel and M Kremer (2004), 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,' Econometrica, Vol. 72 (1): 159-218.
10 Bobonis, Gustavo, Edward Miguel and Charu Puri Sharma (2004), 'Iron Deficiency Anemia and School Participation', mimeo, University of California, Berkeley.
11 Hoyt Bleakley (2004), 'Disease and Development: Evidence from Hookworm Eradication in the American South', mimeo, University of California, San Diego.
12 Abhijit Banerjee, Shawn Cole, Esther Duflo and Leigh Linden (2004), 'Remedying Education: Evidence from Two Randomised Experiments in India', mimeo, MIT.
13 For an example of using a structural approach to analyse data from a randomised experiment, see: Orazio Attanasio, Costas Meghir and Ana Santiago (2003), 'Education Choices in Mexico: Using a Structural Model and a Randomised Experiment to Evaluate PROGRESA', mimeo, The Institute for Fiscal Studies.
14 For a version of an approach that compares structural estimates with a randomised evaluation, see: Petra Todd and Kenneth Wolpin (2003), 'Using a Social Experiment to Validate a Dynamic Behavioural Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Programme in Mexico', mimeo, University of Pennsylvania. They however estimate the structural model without using the experimental data, and then use the experimental data just to confirm their structural results. The problem with this is that the non-experimental estimates may be hopelessly biased but sufficiently ill-estimated to still be consistent with the experimental results. It seems to make more sense to start from the experimental results.
15 Marianne Bertrand, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, Jon Zinman, 'Pricing Psychology: A Field Experiment in the Consumer Credit Market', mimeo, University of Chicago (2005).
16 Dean S Karlan and Jonathan Zinman, 2005. 'Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment', Working papers 911, Economic Growth Centre, Yale University.
17 Esther Duflo and Michael Kremer and Jon Robinson (2004), 'Understanding Technology Adoption: Fertiliser in Western Kenya, Preliminary Results from Field Experiments', mimeo, Poverty Action Lab, MIT.
18 For a discussion of why none of the existing theories help a lot in understanding the Duflo-Kremer-Robinson results on fertiliser adoption, and the challenge it poses for development thinking, see Esther Duflo (2003), 'Poor but Rational?' forthcoming in Abhijit Banerjee, Roland Benabou and Dilip Mookherjee (eds), *Understanding Poverty,* Oxford

University Press.
19 Michael Kremer, Ted Miguel and Rebecca Thornton (2004), 'Incentives to Learn', NBER WP 10971
20 Daniel Chen and Michael Kremer, 'An Interim Report on a Teacher-Attendance Programme in Kenya', mimeo, Harvard University.