Causality and the Experimental Turn in Development Economics

By Jigar Bhatt*

"The most important element of the experimental approach may lie in the power…to vary individual elements of the treatment in a way that helps us answer conceptual questions…that could never be reliably answered in any other way"

--Abhijit Banerjee and Esther Duflo

"It is impossible to be truly problem driven and at the same time committed to a certain method"

--Bent Flyvberg

Over the last decade, there has been an explosion of a particular type of methodology – the randomized controlled trial (RCT) – across the Global South to answer questions in development economics and effectiveness. A 'new guard' of development economists has enthusiastically embraced RCTs for their proclaimed ability to establish internally valid causal effects. Presenting a potential paradigm shift in the field, RCTs have sparked a major debate over the state of theory and methodology in development economics. This paper opens by describing the recent and rapid rise of experiments using RCTs in development economics. It then surveys the debate in development economics over RCTs and evaluates the ways RCTs advance development economics methodologically. It finds that the paradigm shift and future of development economics rests almost entirely on the RCTs' claims to internally valid estimates of causal effects. Other longstanding empirical and theoretical shortcomings in development economics would remain. The paper concludes that existing methods in development economics, RCTs or otherwise, are inadequate for understanding poverty because they fail to address issues of inequality and power. The debates over RCTs, and growth and poverty more generally, have been too narrowly confined to development economists and should be opened to include other social scientists and philosophers. Going forward, a plurality of disciplines and methods should be marshaled to understand development effectiveness.

In his landmark book *The Structure of Scientific Revolutions* Thomas Kuhn describes a paradigm as 'the shared commitments and elements of a scientific group' and these shared commitments as being a 'matrix of various ordered elements'. He describes a paradigm shift as, "A reconstruction of the field from new fundamentals, reconstruction that changes some of

the field's most elementary theoretical generalizations as well as many of its paradigm methods and applications" (Kuhn, 1996).

Economics underwent its first major paradigm shift in the late 1970s. Neo-classical economics came to prominence based on the rigor of its hypothetico-deductive method and assumptions about equilibrium and rational choice within a market economy. Despite its deductive rigor, the general equilibrium (GE) model[1] eventually faced a crisis of legitimacy. Other social scientists resented the model's narrowness in social thought and preference for prediction over explanation. This led to a shift in priorities among the economics' core from formal mathematical models and the hypothetico-deductive approach to a more inductive, empirical approach (Milberg, 2009; Blaug, 1984).

During this transition, explanation of the overall economy through the GE model was being abandoned for practical policy models that attempted to consistently explain partial aspects of the economy such as international markets for goods and domestic markets for labor. Development economists for their part began conducting cross-country growth regressions to identify the successful determinants of economic growth (Cohen, 2008). Rather than pontificating about growth through abstract models and assumptions divorced from reality, development economists began examining the data to find growth determinants in the developing world based on theory developed from successful growth in the industrialized world. Most of this analysis was aimed at the development interventions *du jour* – 'big push' investments that aimed to jumpstart growth in the Western Empire's former colonies (Cohen, 2009).

Later during the 1990s economics' practical policy models also faced a crisis of confidence. Although economists' methods were inductive and their models were more grounded in reality, the models were still considered too theoretical. Policymakers grew frustrated with the models' inoperability and failure to make accurate predictions. From this second crisis emerged the 'empirical turn' in economics in the late 1990s (Milberg, 2009; Mookherjee, 2005). In the place of an emphasis on practical policy models there emerged an emphasis on the sophisticated measurement of variables and statistical techniques applied to extensive quantitative data sets (Milberg, 2009; Mookherjee, 2005). Theory was supplanted by empirics. Economists began to approach traditional notions of scarcity, price, utility, and maximization inductively through robust statistical methods rather than deductively through models with stated assumptions. The paradigm shifted away from deduction and stylized models towards pragmatic induction and statistics (Gudeman, 2009).

Among economic sub-fields, development economics appears to have most adamantly embraced the empirical turn in economics with its reliance on planned experiments and freshly collected data. A 'new guard' is using the RCT, a methodology borrowed from Western biomedicine, to challenge the 'established guard' that has heretofore relied heavily on econometric, or regression, analyses of existing datasets. Some development economists began using RCTs, in principle at least, because of an increasingly polarized debate over disagreement of 'what works' in development and the role development aid plays in poverty alleviation. The new guard is not wedded to existing economic development theory and is open to their experiments challenging existing theoretical assumptions. The enthusiastic

---

[1] The notion that a fully decentralized, competitive private ownership system with self interested and rational actors leads to a socially optimal economic outcome.

adoption of RCTs among an increasing number of development economists is resulting in the 'experimental turn' – arguably the apex of the empirical turn – in development economics.

1.   The Rise of Experiments in Development Economics

The RCT is a particular type of epistemic method that estimates the effect of a treatment, such as a development program or policy, on a treatment group. It does so by comparing the treatment's effect on a treatment group to that of a control group, or counterfactual, that does not receive the treatment. The treatment and control groups are established through a process of random selection that aims to eliminate selection bias. It is an attractive research method because by eliminating selection bias[2] an RCT allows researchers to make confident claims about the treatment's causal effect, within certain limitations and qualifications.

Randomized controlled trials have a rich history in Western biomedicine. The first RCT was conducted by a British doctor named James Lind in 1747. During his early years as a British naval physician, Lind was called upon to treat sailors at sea who were thought to have scurvy. He divided 12 sailors on a ship into six groups of two men each and varied the treatment type administered to each group; one group received oranges and limes while others received cider, sea water, vinegar, nutmeg, or an "elixir of vitriol." The group receiving the citrus fruit recovered the most quickly (Thomas, 1997). Dr. Lind's genius at the time was in his decision to administer unique treatments across the sailor groups[3] (i.e. vary the treatment).

The modern RCT design, widely used in clinical trials for the last 60 years, was developed in 1948, 200 years after Lind's trial, by the British statistician Austin Hill. During a statistical trial on the antibiotic streptomycin, Hill was able to "crack" the statistical design that allowed for statistically significant statements to be made about a treatment's effect on a treatment group. It has become the model for the modern RCT (Thomas, 1997). After the modern RCTs' steady adoption in medical trials in the West, it also began to make its way into social policy. Starting in the 1970s, RCTs became a mainstay in British and American educational and poverty research (Shadish, 2009). Randomized controlled trials have been most popular in industrialized Anglophone countries: the United States, Britain, Canada, Australia, and New Zealand have all incorporated RCTs into research on medicine and policy. Until recently, RCTs in other countries have been rare.

Over the last decade, there has been an explosion of RCTs to examine poverty alleviation interventions' effectiveness in the Global South. At the heart of this phenomenon is the Massachusetts Institute of Technology's (MIT) Abdul-Jamaal Lateef Poverty Action Lab (J-PAL). J-PAL describes itself as "A network of 53 affiliated professors around the world who are united by their use of Randomized Evaluations to answer questions critical to poverty alleviation. J-PAL's mission is to reduce poverty by ensuring that policy is based on scientific evidence."[4] The Lab's founding director, Esther Duflo, an MIT trained economist who in the last two years, among other honors, was named France's top intellectual, won the John Bates

---

[2] The possibility that an outcome was the result of something other than the treatment itself.

[3] For an interesting discussion of the Lind trial see: Bartholomew, M. 2002. *James Lind's Treatise of the Scurvy (1753).* Postgraduate Medical Journal. 78: 695-696.

[4] http://www.povertyactionlab.org/about-j-pal

Clark Medal[5], received a MacArthur Foundation Genius Grant, and was named among both Foreign Policy and Time Magazines' top 100 influential people[6].

Since its founding in 2003, J-PAL has initiated 265 RCTs on five continents. This amounts to an average of three trials per month over the last eight year[7]s. It has spurred partners committed to its mission, namely Innovations for Poverty Action (IPA) at Yale University and the Center of Evaluation for Global Action at the University of California Berkeley. All three research centers regularly initiate and collaborate on their own research and training while also providing consulting services to the major development and philanthropic institutions such as the World Bank and Bill and Melinda Gates Foundation (BMGF).

The majority of J-PAL's evaluations (quantities in parentheses) take place in Africa (79), South Asia (65), and Latin America (47). The most common policy goals trials attempt to answer questions in are: access to credit (42), community participation (23), educational quality/student attendance (31/19), and the encouragement of savings (26). The themes, or sectors, covered by J-PAL include: finance and microfinance (93), education (59), health (54), political economy and governance (42), labor markets (20), agriculture (16), and environment and energy (13)[8]. There is no explanation on J-PAL's website why these policy goals or themes were chosen.

Randomized controlled trials have had a major impact on the practices of multilateral development institutions, such as the World Bank, major philanthropic foundations such as the BMGF, nongovernmental organizations involved in poverty alleviation and development, and Ministries in developing countries. Scholarly work on economic development has had a long history of affecting economic development policy (Hagen, 1988), however, the speed with which J-PAL has transformed the way major institutions do research and evaluation is noteworthy. The Millennium Challenge Corporation, a new federal U.S. foreign aid initiative focused on results and development effectiveness based its entire multi-million dollar evaluation strategy on the J-PAL method and the United States Agency for International Development is considering doing the same (Shah, 2010).

RCTs have many adherents, particularly among economists and quantitative evaluators. In the development effectiveness field, a handful of increasingly visible champions in America's top economics departments are endorsing RCTs. These economists use RCTs as a micro-approach to the hotly contested macro issue of aid effectiveness. Development aid's chief proponent, Jeffrey Sachs, argues in *The End of Poverty* that a surge of development financing is necessary for countries to overcome the 'bottom rung' and enter the global market economy (Sachs, 2005). Critics, most notably William Easterly and Dambisa Moyo, suggest that current aid practices are the problem and perpetuate countries' poverty; additional aid will either do no good or potentially cause further harm (Easterly, 2006; Moyo, 2009). Duflo considers both these arguments to be extreme and not based on evidence. She believes that we are able to have such widely held views on aid effectiveness and impact on poverty

---

[5]An annual award by the American Economics Association to the economist under 40 that the association believes has made the greatest contribution to the field of economics.

[6] http://econ-www.mit.edu/faculty/eduflo/short

[7] http://www.povertyactionlab.org/search/apachesolr_search?filters=type:evaluation

[8]http://www.povertyactionlab.org/search/apachesolr_search?filters=type:evaluation. The regions, themes and policy goals are not mutually exclusive. Total number of evaluations equal 265.

because we lack a counterfactual, "Africa got a lot of aid [between 1960 and 2006] but GDP is not making much progress, but do you know what would have happened without the aid? GDP might have been much worse, or it could have been much better, we don't know what the counterfactual is – *there is only one Africa*" (Duflo, 2010). The message is that with a well-established counterfactual and evidence on outcomes we can resolve the aid effectiveness debate "from the ground up" one experiment at a time.

2.   *The Debate over the Experimental Turn*

The debate around RCTs' value in development effectiveness has taken place almost exclusively among development economists. The RCT proponents, or new guard, have enthusiastically advocated their approach while the established guard has cried foul (Deaton, 2009; Mookherjee, 2005; Bardhan, 2005; Ravallion, 2009; Basu, 2005). To understand the debate surrounding this paradigm shift one needs to look at how the new guard is claiming to advance the field of development economics. The new guard's claim is not to have found some new theoretical facts[9]. They do not claim to have discovered some new determinant of growth, such as a particular policy, program, technology, or institutional innovation. In fact, quite the opposite – the new guard eschews theory and claims that the existing growth and development theories prevent learning and progress within the discipline (Banerjee, 2008). In keeping with the empirical turn in economics, they claim that their methodological advancement in the field will allow for internally valid, causal statements that will provide policy makers the evidence they need, something previous econometric studies and practical policy models failed to do (Milberg, 2009; Cohen, 2008). These confident causal statements are thought to advance the field of development economics. This unsettles many development economists, particularly those employing econometric techniques, who believe their work is based upon a theory of causal mechanisms (Deaton; 2009; Mookherjee, 2005).

The new guard claims that its approach is an improvement over existing methods in development economics. The established guard has focused its critiques around issues of RCTs' estimates' validity and avoided a broader critique of econometrics. This is most likely because econometrics, the established guard's primary research approach, has itself faced a number of critiques as an investigative method in economics research. However, neglecting to understand RCTs in relation to econometrics has limited our understanding of the experimental turn.

Researchers investigating econometrics have found that it suffers from: i) poor consistency of measurements; ii) failure to replicate studies for scientific verification (i.e. high Birge ratios); iii) inadequacy of conventions for assessing empirical results; iv) mistaking statistical significance for economic significance; v) transparency (e.g. 'data mining'); and vi) failure to establish causality. For RCTs to represent a true progression in development economics' methodology they must overcome the limitations inherent in econometrics. We now take up

---

[9]New theoretical and empirical facts are necessary conditions for a scientific program to be 'progressive' in a Lakatosian sense. Imre Lakatos, a philosopher of science, describes scientific progress in terms of developing new facts – statements supported by empirical evidence based on a shared set of paradigms by a scientific community – that can be confirmed or falsified through scientific tests.

each of these limitations in turn.

Macro-level economic data, the kind most frequently used in cross-country growth regressions, are fraught with measurement error. Indicators representing macro-level economic conditions such as production through gross domestic product (GDP), inflation through consumer price index (CPI), inequality through Gini coefficients, and trade through the ratio of imports plus exports to GDP (trade-share) often do not measure the same thing(s) across observations (Hollander, 2011; Rodriguez, 2007). For example, the European and American central banks both target inflation, but the American CPI target excludes food and fuel, allowing more lenient policy when oil prices rise, while its European counterpart monitors overall inflation (Hollander, 2011). A related issue is whether the way these variables are measured corresponds to the policy tools available. While policy makers may have control over trade inducing measures such as tariffs, growth regressions often include trade-share as a determinant of growth. Policy makers, however, have much less control over trade shares than they do over tariffs (Rodriguez, 2007).

Experiments claim to get around these regression measurement challenges by tailoring data collection to a 'specific measurement problem': data is typically collected especially for the purpose of the experiment (Banerjee, 2008). This can help ensure that economic conditions are measured with the policy makers' options in mind. It does not solve the challenge, however, of being able to compare across experiments unless inputs and outcomes are monetized. This process also comes with its own set of drawbacks.

Birge ratios, the relative magnitude of the internal and external consistency of measurements in a specific field, are orders of magnitude higher in economics than of those generally found in physics and psychology[10] (Mirowski, 2004). The Birge ratio (in parentheses) is a meta-analysis of published estimates, such as the speed of light in physics (1.42), spatial perception in psychology (1.64), and the United States employment-output elasticity in economics (22.70). The close proximity of Birge ratios between psychology and physics makes it more difficult to dismiss the large Birge ratios in economics as resulting from inherent differences between the natural and social sciences. It makes the case for reliable econometric analyses that much more difficult. According to proclamations by the published authors, the econometric based estimate closely simulated controlled experiments (Mirowski, 2004). However, it is subject to debate whether an econometric model can truly simulate a controlled experiment and the RCT proponents might argue that Birge ratios could be improved through RCT produced estimates. Currently, too few estimates have been produced and replicated by RCTs to determine if the methodology can improve economics Birge ratios.

One of the ideals of science is replication of investigations. Economists frequently reproduce but rarely replicate (Mirowski, 1991; Summers, 2005). That is, they continue to analyze the different determinants of growth, but rarely will economists try to exactly replicate an existing econometric investigation of growth step by step. This is due to the high monetary, logistical, and professional costs of doing so[11] (Mirowski, 1991). According to this

---

[10]The larger the Birge ratio, the less successful the measurement is at taming error.

[11]Some challenges include whether to simply reprocess the data or recollect and collate the data and whether to replicate an analysis, which is professionally undesirable, or produce an original analysis. These are in addition to the philosophical question of whether an analysis can ever truly be reproduced, for the simple reason that time has passed and

assessment, there is reason to believe that the replication issue applies equally to econometric analyses and experiments; RCTs will face the same challenges in replicating their results.

Another issue of empirical evidence from econometric studies is what William Milberg (2009) calls the 'Kletzer effect'. It turns out that a given econometric result can support various theories or be given radically different interpretations. A study by economist Lori Kletzer in 2001 is used to support both sides in the debate over the effects of rising international outsourcing by U.S. firms, one side that claims international trade benefits US workers, the other arguing that it is harmful to the interests of American workers (Milberg, 2009). Because the Kletzer effect is the result of economics not being able to adequately establish conventions for assessing the merits of empirical results, it applies to econometric and RCT based estimates equally.

Dierdre McCloskey has assailed the misuse of statistical significance by economists in regressions. She and coauthors have taken particular aim at economists' assertions that statistical significance equates to economic significance and have also documented poor reporting of significance tests' results in published articles (McCloskey, 1996). Drawing on McCloskey's analysis there is no reason to believe that RCTs are immune from the temptation to report statistical significance as economic significance or shoddy reporting of experimental processes and tests' results. In fact, one study suggests that careless reporting of RCTs' results is already a problem (Bruhn, 2009).

Economists have raised the issue of transparency in econometric analyses since the early 1980s (Leamer, 1983). The most common problem has been data mining – running numerous econometric model specifications until a statistically significant result is achieved (Leamer, 1983; McCloskey, 1996; Duflo, 2006). For example, in a time series analysis, one may not get significant results for a 30 year period but will for a 20 year period. Although the researcher set out to investigate the political economy of a 30 year period for hypothesis testing, if the data show a significant effect on the dependent variable for a 20 year peiod, s/he may "adjust" her hypothesis to match her findings. A researcher would likely do this because of publication bias – journals are more interested in publishing positive or significant results.

Researchers conducting experiments claim that RCTs lead to greater transparency and simplicity (Banerjee, 2008; Duflo, 2006) and potentially resolve the data mining issue. Because of the objectivity of the RCT method and certainty of the experiments' results, experimenters cannot just choose the method that leads to the statistically significant or most positive result. In an econometric based evaluation, for example, different model specifications can be run until a statistically significant result is achieved (Duflo, 2006). Because of the subordinate relationship an evaluator has with a project or program implementer, there may be an incentive to publish results favorable to the implementer. With RCTs, however, there is only one model and therefore only one result (Duflo, 2006).

RCTs hold the potential to improve the transparency of development economics' empirical work but they cannot entirely resolve the issue. A recent paper assessing the quality of published RCTs (Bruhn, 2009) found that many researchers were not transparent about the way they conducted random assignment. Upon closer investigation through field work and interviews with the researchers, the authors found that in many cases results were reported as if they were conducted in randomized trials but in fact they were derived from matching

---

the new analysis conducted in *t* will be taking place in *t*+1 (Mirowski, 1991).

exercises[12]. Transparency around the randomization process is important because it determines if the treatment caused the effect; non-random selection methods, like matching, increase bias and cast doubt over causal effects (Leamer, 1983). There is also no reason to believe that researchers using experiments will not be affected by publication bias to have statistically significant results published over statistically insignificant ones. In short, there is nothing inherent to the RCT methodology that can eliminate the other transparency challenges that affect empirical work in economics more generally.

It appears that RCTs are unable to fully resolve the methodological issues that have haunted econometric analyses. There exists some potential for improvements in measurement and transparency. However, failure to replicate studies for scientific verification, inadequacy of conventions for assessing empirical results, and mistaking statistical significance for economic significance will all continue to be a problem with RCTs in development. If RCTs cannot resolve these several longstanding issues that have created a crisis in confidence around econometric work, why are they generating so much change and debate, potentially leading to a paradigm shift in the discipline?

3.   *The Debate over the Experimental Turn*

To understand the RCT's allure in development economics we must turn our attention to causality. One of the fundamental critiques of econometrics is that adequate instruments cannot be found, and as a result causality in an econometric analysis is rarely, if ever, achieved. This has been a frustration of the cross-country growth regression literature (Cohen, 2009). The RCT's allure is not, as we have already discussed, a result of its ability to overcome the many longstanding issues that have generally plagued empirical work in econometrics. Instead, its power and draw comes from its seeming ability to overcome just a single one – the inability to confidently establish causal relations between independent and dependent variables. We can understand how the RCT moves beyond the empirical limitations of econometrics by demonstrating how it methodologically advances the workhorse of econometrics, the simple linear regression model.

In a simple linear regression model (LGM) the dependent Y variable, for example a continuous variable such as students' test scores, is estimated using a number of independent X variables that the researcher believes has some determining effect on the dependent variable. The model might look something like this: $Y = \alpha + \beta x + \epsilon$. Here, α is the constant, βx are the set of independent observable variables, or determinants, that the researcher has chosen because s/he believes they determine test scores. The unobserved error term, ϵ, is assumed to be a random variable (in actuality, infinite variables) with specified properties unique of the observable determinants (Ramanathan, 1995). In other words, the error term is all the complexity not captured by the researcher's regression model[13].

---

[12]Matching is a close approximation to a RCT (Bruhn, 2009) but not any more advantageous than econometric work with a good instrument or a natural experiment. Describe matching in some detail

[13]Not all regression models are linear. Depending on the character of the dependent and independent variables, regressions can be non-linear (e.g. exponential, logarithmic), polynomial,

RCTs, by contrast, reduce the misspecification uncertainty to zero (Leamer, 1983). This statistical condition can be translated to mean that bias from the model has been removed. The randomization process of assigning units to the treatment and control groups 'does the work' of eliminating the error term, or the uncertainty inherent in the LGM. This is achieved through the rigorous establishment (i.e. through random assignment) of a counterfactual. A properly implemented RCT with a positive result might allow a researcher to make the following claim: "Compared to the control group, the introduction of school meals in the treatment school increased the treatment school's students' math test scores by a statistically significant 7 percent". The key advantage of experimentation is through the administration of the treatment after the randomization process is complete (Imai, 2010). This, advocates say, eliminates the likelihood of something other than the treatment, such as a new textbook program, affecting the treatment's outcome (e.g. test scores). In a simplified linear regression investigation such a claim is more difficult to make. To make an equal claim, researchers must include every possible independent variable that might affect test scores – a challenging, if not impossible, endeavor.

These causal claims allow for something statisticians call internal validity. According to Banerjee (2008), effective policymaking requires making judgments about the efficacy of individual components of programs, "Without much guidance from *a priori* knowledge… it is also difficult to learn about these individual components from observational (i.e. nonexperimental) data" (Banerjee, 2008). By varying one factor of intervention at a time, RCTs are able to come to internally valid estimates of the causal effect of different alternatives (Banerjee, 2008)[14]. This approach then gives policymakers the range of intervention options and their respective costs and benefits. Thus, RCTs provide internally valid estimates of benefits compared to the costs of various alternatives; policymakers presumably can then choose the most efficient options.

With both the new and old guards conveniently dismissing the majority of econometrics' limitations, the debate over methodological advancement (and a paradigm shift more generally) has hinged on RCT estimates' validity. The established guard has retorted that despite the RCT's elimination of bias the methodology is not without its tradeoffs. They argue that RCTs are too dependent on their environment, in other words, they may have internal validity but they suffer from the lack of external validity. Thus, there is no guarantee that a successful outcome as measured by the RCT will have a similar effect if implemented elsewhere. There are two components to the external validity issue. First, there is concern as to whether the treatment can be translated to other environments.[15] Second, there is concern as to whether the treatment effect reflects the unique characteristics of the implementer: the

---

or logistic (when the dependent variable is a binomial). These are all, however, variations on the ordinary least squares linear regression model and bounded together by the need for an error term.

[14]For example, to improve attendance at school should education providers reduce class size by adding teachers, increase access to textbooks, or provide school lunches? What does each cost and what is the resulting attendance increase from each?

[15]e.g. the kids at the treatment village's schools attendance went up by one day per month but there is no guarantee that children's attendance at a village in a far away district will also go up.

treatment effect may differ when the implementer is different[16] (Banerjee, 2008, Deaton, 2009).

Similarly, there is criticism surrounding what economists call 'equilibrium effects'. Program effects found in a small study may not generalize when the program is scaled up, for example, nationwide. As one of the goals of RCTs is to find interventions that work, it is likely that implementers looking for solutions will want to replicate the intervention. Critics argue that RCTs cannot assure successful replication of positive effects (Deaton, 2009). The advocates recognize these limitations and suggest that larger RCTs be conducted over smaller RCTs and that non-experimental studies complement experimental ones (Banerjee, 2008).

RCTs also suffer from homogeneity of treatment effects. One notable critic has targeted RCTs for only being able to provide the mean effect of a treatment (Deaton, 2009). Heterogeneity of treatment effects, across sub-groups of an important demographic characteristic, for example, cannot be reported with statistical significance. RCT advocates confess that RCTs cannot address the mobility of the population (Banerjee, 2008). This is a serious shortcoming of the RCT because it prevents it from addressing issues of distribution and inequality.

RCTs are unable to address many of the challenges facing empirical work in economics more broadly and the new guard rests its reputation almost entirely on RCTs' ability to provide internally valid estimates of causal effects. However, as the established guard has pointed out, internally valid results are neither necessarily externally valid nor maintained in the face of program or policy expansion. More importantly, some doubt if field based experiments really are objective and scientific. Leamer (1983) says, "[t]he fact of the matter is that no one has ever designed an experiment that is free of bias, and no one can" and Friedman (1953) tells us "[t]he difference between scientific experiments and natural experiments is difference in degree, but not in kind." Since natural experiments' effects can be and are estimated econometrically, the RCT in essence is an extension of the LGM that has the power to establish a particular type of causality. It might then be appropriate to suggest that the future of development economics research rests on the role causality plays in development economists' "disciplinary matrix."

This debate over experiments in development economics is not one that will be resolved easily, "The question about the relationship… between causes and correlations is one of the crucial questions of methodology of econometrics" (Moneta, 2011). Herbert Simon's approach to causality in econometrics established the contemporary argument for experimental evidence to establish causality, particularly when an adequate instrument was not available for an econometric regression[17] (Hoover, 2008). Going back further, John Stuart Mill provided the philosophical foundations for efficient causality (also known as productive

---

[16]e.g. the local department of education in district X has a high capacity to procure textbooks but the same textbooks may not have an equally positive effect on attendance in district Y because the capacity of its local department is lacking.

[17]Herbert Simon was one of the twentieth century's most influential social scientists. This experimental approach emerged from his participation in the Cowles Commission at Yale University in the 1940s; the Cowles Commission's goal was to link economic theory to mathematics and statistics (http://cowles.econ.yale.edu/P/reports/1932-52.htm#History).

causality) and the experimental method through his example of the method of difference. However, unlike Herbert Simon, Mill was skeptical as to whether such a method could be used to establish causality in the "moral sciences": "An *experimentum crucis...* converts the presumption we had before of the existence of a connection between A and B into proof, by negating every other hypothesis which would account for the appearances. But this can seldom be done in the moral sciences, owing to the immense multitude of the influencing circumstances, and our very scanty means of varying the experiment" (Mill, 1836).

4. Experimental Turn or U-Turn?

With a vigorous debate over RCT results' validity being waged between the new and established guards, a paradigm shift in development economics does not appear to be a foregone conclusion. Thus, we have to ask the question, "What if the established guard prevails?" The established guard prides itself on the use of theory and advancement of the discipline through relevant policy models' refinement. However, previous econometric growth models have not enlightened us (Cohen, 2009; Rodriguez, 2007) and the research program on growth regressions is not progressive in terms of providing us with new empirical facts (Cavusoglu, 2006). This has less to do with modeling or data issues and more with the "scientific illusion" (Summers, 1991) in development economics.

In the development economics debate thus far there has been little discussion of the role of theory, except the established guard's claims that the new guard threatens the scientific nature of development economics by abandoning development theory (Deaton 2009; Mookherjee, 2005). One complaint is that RCTs do not test based on economic development theory and therefore do not contribute to its scientific research program (SRP). Thus, they can tell us *whether* a program worked but not *why* a program worked (Deaton, 2009; Mookherjee, 2005; Sayer, 1992). They can tell us about causality, but little to nothing about causal mechanisms – why a particular outcome came about. Meanwhile, RCT advocates suggest that experiments allow a "clean test" of the theory because they are not constrained by pre-existing ideas. They claim this approach has even helped rethink some basic assumptions previously held about neoclassical demand theory and that as more experiments are completed the results will lead to an evidence base and an inductive theory building process (Banerjee, 2008).

By debating the role of theory in development economics both sides are implicitly making two important assumptions: 1) that development economics is a progressive science; and 2) there is such a thing as development theory. A progressive SRP, as outlined by the philosopher Imre Lakatos, requires a field to produce novel theoretical and empirical facts (Lakatos, 1970). Research using the Solovian growth model[18] has been found to be progressive, but Solovian models are also flawed for assuming that technology, a determinant of growth, is considered exogenous – a global public good equally available to anyone.

---

[18]One of development economics principle concerns is finding determinants of growth. One of the pioneers in explaining growth has been the economist Robert Solow, best known for the Solow Growth Model.

Endogenous growth models have attempted to correct this shortcoming by, as their name suggests, making technology an endogenous predictor of growth. Research using endogenous growth models, however, has not been found to be scientifically progressive (Cavusoglu, 2006). Development economists' research program on growth is meaningful but has not been found to be progressive in a Lakatosian sense, suggesting that it does not measure up to more 'scientific' programs in the natural sciences (Cavusoglu, 2006).

The second assumption, that development theory is somehow universal, objective, and therefore confirmable and falsifiable, is a questionable claim. Development, which in economics is primarily concerned with growth in the former colonies of European powers, is highly politicized. The concept of development originated with former President Truman's inaugural speech in 1949 (Rist, 1997). The ideas of Marx and Hegel were abandoned during McCarthyism and development theory has since proceeded ahistorically (Leys, 2005). Theories of a world division of labor resulting from European colonialism had no place in the quest for higher gross domestic product – all players are now assumed to be working with the same system towards the same goals (Rist, 1997).

Since Truman's speech, development theory has had a tendency to shift frequently based on the whims of academics and politicians in the United States[19]. These periodic revisions to development theory make a progressive SRP that creates new theoretical and empirical facts through testable hypotheses improbable. This is in contrast to revolutions in the natural sciences that occur over centuries as opposed to decades (Kuhn, 1996). Development theory's ahistorical and politicized approach that has implicitly barred ideas unfavorable to American interests has contributed to development policy models' and development econometric research's unreliability. The established guard's claims that the new guard's a-theoretical approach is detrimental to the progress of development economics can only be taken seriously if the political, ahistorical, and unscientific nature of existing development theory is also adequately addressed. Until then, critiques that the new guard is threatening the theoretical foundations of development will appear overstated.

The new guard, for its part, purports that it is advancing the discipline of development economics on methodological, rather than theoretical, grounds. They say they are advancing existing econometric methods, but I would argue that the existing methods are inherently insufficient; so this new "advancement" must also be insufficient. To understand the inadequacy of econometric estimates in poverty research, experimental or otherwise, we must explore what these estimates are unable to tell us.

RCT estimates are not well suited to address questions of distribution (Deaton, 2009). This is particularly worrisome since inequality worldwide, despite not receiving mention in J-

---

[19]In the 1950s and into the 1960s a positivist orthodoxy based on the 'big push' idea of state-led development reigned (Leys, 2005; Hagen, 1998). In the 1960s and 1970s modernization theory was presented as an alternative to the failures of the big push and the United States took a less direct role in foreign economic development (Leys, 2005; Amsden, 2007). The 1980s gave birth to the Washington Consensus and market-led development that carried well into the 1990s (Gore, 2000; Amsden, 2007). Currently there is a search to replace the Washington Consensus with some calling for a return to state-led development (UNCTAD, 2009).

PAL's evaluations,[20] has been steadily rising (Chakravorty, 2006; Pieterse, 2002). While incomes have clearly risen in the past decade, evidence shows that inter- and intra-country inequalities have also risen. On a global level, estimates of the income gap between the fifth of the world's people living in the richest country and the fifth in the poorest went from 30:1 in 1960 to 74:1 in 1997 (UNDP, 1999). Certain segments of society have seen considerable welfare gains – raising the global mean or the mean for their countries – while large segments of society have seen no gain and even declines in welfare (Wade, 2004). As an illustrative example of the overall trend, the Gini coefficient in India increased from 0.32 to 0.38 between 1978 and 1997 even while absolute poverty declined[21] (Chakravorty, 2006; p. 56).

Despite progress on the Millennium Development Goals (MDG) to halve extreme poverty and hunger by 2015, promote health and education for all and tackle gender inequality, persistent inequalities exist within countries. The creators' decision to measure MDG progress in terms of "averages" and "proportions" has dealt it the same fate as RCT estimates: the MDG average estimates have occluded distributional aspects of progress across social groups and the outright exclusion of some groups from any benefits based on their racial, ethnic, lingual or religious identities (Kabeer, 2010). Urban inequality is particularly acute. Throughout Sub-Saharan Africa, Latin America, and Asia urban inequality is higher than rural inequality. Also, urban inequality is often higher than overall national inequality[22]. Meanwhile, as inter- and intra-country, ethno-cultural, and urban inequality rises, a growing body of literature on subjective well-being (SBW) is demonstrating that, contrary to many economists' claims, our income relative to our peers' does matter to our own well-being (McBride, 2001; Luttmer, 2005; Ferrer-i-Carbonell, 2005).

Econometric estimates in development economics do not shed light on questions of power relations or global politics. This has been a criticism of poverty research and the poverty reduction agenda in general (Cammack, 2004). By ignoring questions of politics, inequality, neoliberalism, and power, issues most other disciplines such as political science, sociology, anthropology, planning, and geography openly address, mainstream development economics

---

[20]A full-text search of J-PAL's online database of 265 evaluations for the words 'inequality', 'equality', 'inequity', and 'equity' generated a total of 5 hits. A search for 'neoliberalism' returned 0 hits. By contrast, the word 'poverty' generated 112 hits and the word 'poverty line' generated 11 hits.

[21] Even declines in poverty are coming under intense scrutiny. Using updated and improved cost of living data, World Bank, which has long lauded global markets for reducing absolute poverty of those living on less than $1 a day, has had to admit that it overestimated the number of people brought out of poverty by 415 million "In the developing world outside China, the $1.25 poverty rate has fallen from 40 percent to 29 percent over 1981-2005. However, given population growth, this progress was *not enough to bring down the total number of poor outside China, which has stayed at about 1.2 billion"* (World Bank Press Release, August 2008).

[22]Statements based on analysis of Gini coefficients. See UNHABITAT, "State of the World's Cities 2008/2009 – Harmonious Cities". http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=2562

has constructed a tunnel through which we see poverty in terms of "dollars per day."

Power relations are conspicuously absent from the leading dollars-per-day poverty accounts, "What is missing is a problematisation of poverty itself…Economists tend to use culturally flat definitions of poverty, as if monetary income measures hold universal validity" (Pieterse, 2001; p. 1026). On a global scale "[p]overty is a safe theme: the numbers are worrying, but isn't poverty mostly concentrated in distant lands? Has unequal development not been the rule of history, particularly since the Industrial Revolution? ... Economists and the international institutions that employ them routinely ignore differences of power; by prioritising poverty over inequality, relations of power and the responsibilities these entail are eliminated from the picture" (Pieterse, 2001; p. 1027).

The current poverty discourse has prevented us from having real discussions on inequality or the root causes of poverty. Instead, we feel compelled to seek answers with statistical significance on issues with oftentimes obvious answers that are hardly "economically significant": "One of her [Esther Duflo's] biggest findings is that microfinance, the poverty-reduction solution *du jour*, isn't all it's cracked up to be. Which, like many great economic insights, seems obvious when you think about it; after all, not everyone is born to be an entrepreneur" (Foroohar, 2011). Microfinance is much less about entrepreneurship and much more about the financialization of development, a form of "poverty capital" and power (Roy, 2010) in an era of market worship and neoliberal state retraction. This political economic and sociological assessment of microfinance was produced by a planner, not a development economist, and is essential to understanding what some consider to be 'the most important economic phenomena since the advent of capitalism' (Khosla, as cited in Roy, 2010).

The issue for development and poverty more generally is not whether to use regression or experimental estimates, but that the issues and analysis have been dominated by economists, creating a virtual "disciplinary monopoly" in what is far from a "free market of ideas" (Rao, 2007) in an area where no single discipline can have "the" answer. Development economists have an important role to play in research on growth, poverty, and inequality, but they do not have monopoly rights to research on poverty or growth nor do they have a monopoly on establishing causality. Causal explanation has been the concern of philosophers in the West since the Scottish Enlightenment and every field of social science engages in causal explanation, even if it is not of the efficient variety (Vayda, 2011).

There needs to be more than a partial advancement of an econometric method through the establishment of efficient causality if we are to tackle the enormous challenges of poverty and inequality. We need to engage a plurality of voices on issues of poverty (Rao, 2007), understand what institutions like the World Bank mean when they say "poverty reduction" (Cammack, 2004; Roy, 2010), bring global politics and inequality back into discussions of poverty (Pieterse, 2002), and implement methods that explicitly take issues of power into account (Flyvberg, 2004).

5. *Conclusion*

The new guard has opened up a fissure in development economics but the debate thus far has been too focused on the credibility of a supposed methodological improvement. Both the

established and new guards misinterpret the role of theory in development economics. Given the nature of the debate thus far, it appears that neither guard will be forced to confront history, the unscientific nature of their discipline, or the monopolistic role they have come to hold on poverty research. Methodological advancements like the RCT, as well intentioned as they are, will only serve to advance one of several limitations in development economics research. If there is in fact a paradigm shift, an experimental turn in favor of RCTs, it will appear that development economics has indeed become scientific by adopting natural science's most cherished research tool. This research tool, however, is itself coming under increasing scrutiny, potentially placing the new development economics on as shaky ground as its predecessor.

RCTs are an epistemic method that philosophers of science have been debating since their use accelerated in medicine in the 1970s (Cartwright, 2010). Philosophers have considerable doubt as to whether natural science methods developed for laboratories and clinical settings such as RCTs can be translated to non-laboratory settings (Batie, 2008; Biesta, 2010; Cartwright, 2010). More disconcerting perhaps for the new guard is skepticism of RCTs in biomedicine. As one physician and researcher who studies health in developing countries put it "Randomized trials are not infallible, as medicine's sordid history demonstrates. Economics could learn from that experience" (Bendavid, 2011). Still, the nature of medical research allows for greater standardization and thus quality and generalizability, but the same cannot necessarily be said for development economics "Using research outcomes in practice is acceptable in medicine only when the patient-group definition is precise, the disease is similar among patients, and the treatment is equivalent across settings. This is not true for development field trials…this makes generalization outside the specific setting and intervention a near impossibility" (Bendavid, 2011).

If we are to be truly 'problem driven' towards finding solutions to poverty we cannot wed ourselves to a certain method (Flyvberg, 2004). Given development economists ongoing domination of poverty research and the new guard's commitment to experimental method, we are witnessing a new methodological hammer with which questions of poverty are being pounded.[23] The debate on RCTs in development effectiveness may be taking place in the field of development economics, but it cannot be left to development economists alone to settle. Issues of development effectiveness, poverty, and causality are of equal concern to many fields. These other fields have meaningful contributions to make that have thus far been drowned out by the din of efficient causality and monetary income measures. It is time that the voices of other social scientists and philosophers are amplified on the debate around the experimental turn in development economics.

---

[23]As the fabled 'Law of the instrument' or Maslow's Hammer goes, 'Give a boy a hammer and everything he encounters needs pounding' (Maslow, 1966).

REFERENCES

Amsden, Alice H. 2007. *Escape from Empire: the Developing World's Journey through Heaven and Hell.* Cambridge, MA: MIT.

Banerjee, Abhijit, and Esther Duflo. 2008. "The Experimental Approach to Development Economics." NBER Working Paper 14467, National Bureau of Economic Research.

Batie, Sandra S. 2008. "Wicked Problems and Applied Economics." American Journal of Agricultural Economics 90(5): 1176-1191.

Blaug, Mark. 1984. *The Methodology of Economics: or How Economists Explain.* Cambridge, MA: Cambridge UP.

Cammack, Paul. 2004. "What the World Bank Means by Poverty Reduction, and Why It Matters.." *American Journal of Agricultural Economics* 90(5): 1176-1191.

Cartwright, Nancy. 2010. "What Are Randomised Controlled Trials Good For?" *Philosophical Studies* 147: 59-70.

Cavusoglu, Nevin and Tebaldi, Edinaldo. (2006). "Evaluating Growth Theories and their. Empirical Support: An Assessment of the Convergence Hypothesis." *Journal of Economic Methodology* 13(1), 49-75.

Cohen, Jessica, and William Easterly (2009). *What Works in Development?: Thinking Big and Thinking Small.* Washington, D.C.: Brookings Institution.

Deaton, Angus. (2009) "Instruments of Development: Randomization in the tropics and the search for the elusive keys to economic development". NBER Working paper 14690, National Bureau of Economic Research.

Duflo, Esther. "Social Experiments to Fight Poverty". TED Talk, May 3, 2010. Access online: http://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty.html

Duflo, Esther and Michael Kremer. (2005) "Use of Randomization in the Evaluation of Development Effectiveness". In: *Evaluating Development Effectiveness. World Bank Evaluation Series.* New Brunswick: Transaction Publishers.

Easterly, William. (2006) *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done so Much Ill and so Little Good.* New York: Penguin.

Flyvbjerg, Bent. (2004). "Phronetic Planning Research: Theoretical and Methodological Reflections." *Planning Theory & Practice* 5(3): 283-306.

Foroohar, Rana. (2011). "Esther Duflo - The 2011 TIME 100" Time Magazine (16 May 2011).

Friedman, Milton (1953). "Essays in Positive Economics", In: *The Philosophy of Economics: An Anthology*, Hausman, Daniel (ed.), Cambridge, UK: Cambridge University Press.

Gore, Charles. (2000). "The Rise and Fall of the Washington Consensus as a Paradigm for Developing Countries." World Development 28(5): 789-804.

Gudeman, Stephen. (2009). *Economic Persuasions.* New York: Berghahn.

Hagen, James M., and Vernon W. Ruttan. (1988). "Development Policy under Eisenhower and Kennedy." *The Journal of Developing Areas* 23(1): 1-30.

Hoover, Kevin D. (2011). "Causality in Economics and Econometrics." The New Palgrave

Dictionary of Economics, 2nd Edition.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. (2010). "Experimental Designs for Identifying Causal Mechanisms." Princeton University.

Milberg, William. (2009). "The New Social Science Imperialism and the Problem of Knowledge in Contemporary Economics." in Economic Persuasions, Gudeman, Stephen (ed.) Oxford: Berghan.

Mill, John Stuart (1836). "On the definition and method of political economy", In: *The Philosophy of Economics: An Anthology*, Hausman, Daniel (ed.), Cambridge, UK: Cambridge University Press.

Mookherjee, Dilip. (2005). "Is There Too Little Theory in Development Economics Today?" *Economic and Political Weekly* 40(40).

Kuhn, Thomas S. (1996). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago.

Lakatos, I. and A. Musgrave (ed.). (1970). *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.

Leamer, Edward. (1983). "Let's Take the Con Out of Econometrics." American Economic Review 73(1): 31-43.

Leys, Colin (2005). "The Rise and Fall of Development Theory." The Anthropology of Development and Globalization: from Classical Political Economy to Contemporary Neoliberalism. Malden, MA: Blackwell Pub.

McCloskey, Dierdre, and Stephen Ziliak. (1996) "The Standard Error of Regressions." The Journal of Economic Literature 34: 97-114.

Bruhm, Miriam and McKenzie David. (2009). "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." American Economic Journal: Applied Economics 1(4): 200-32.

Mirowski, Philip. (2004). *The Effortless Economy of Science?* Durham: Duke UP.

Moyo, Dambisa. (2008). *Dead Aid: Why Aid Is Not Working and How There Is Another Way for Africa*. London: Allen Lane.

Pieterse, Jan Nederveen. (2002) "Global Inequality: Bringing Politics Back" *Third World Quarterly* 23(6) : 1023-046.

Ramanathan, Ramu. (1995). *Introductory Econometrics*, Fort Worth: Dryden.

Rao, Vijayendra, and Michael Woolcock. (2007). "The Disciplinary Monopoly in Development Research at the World Bank." *Global Governance* 13: 479-84.

Rist, Gilbert. (1997). *The History of Development: from Western Origins to Global Faith*. London: Zed.

Roy, Ananya. (2010). *Poverty Capital: Microfinance and the Making of Development.* New York: Routledge.

Sachs, Jeffrey. (2005). *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin.

Shadish, William R., and Thomas D. Cook. (2009). "The Renaissance of Field Experimentation in Evaluating Interventions." *Annual Review of Psychology* 60(1): 607-29.

Summers, Larry. (1991). "The Scientific Illusion in Empirical Macroeconomics." *The Scandinavian Journal of Economics* 93(2).

Thomas, Duncan P. (1997). "Sailors, Scurvy, and Science." *Journal of the Royal Society of Medicine* 90: 50-54.

Vayda, Andrew Peter., and Bradley B. Walters. (2011). *Causal Explanation for Social Scientists: a Reader.* Lanham, MD: AltaMira.